

Access DB#

32246

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: MICHAEL PAV Examiner #: 72476 Date: \_\_\_\_\_  
 Art Unit: 1646 Phone Number 305-7088 Serial Number: 09/2000 548  
 Mail Box and Bldg/Room Location: 611/10813 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Novel human cytokine / chemokine

Inventors (please provide full names): CURVA GULI

Earliest Priority Filing Date: \_\_\_\_\_

\*For Sequence Searches, provide the appropriate serial number.

\_\_\_\_\_ (others) along with the

(DIV of 08/822,264)

Please search SEQ ID NO: 1-2

Interferon and Chemokine Database

Reverse Transcriptase SEQ ID NO: 1

Thats

Point of Contact:  
 Beverly Shears  
 Technical Info. Specialist  
 CM1 12C14 Tel: 308-4994

## STAFF USE ONLY

## Type of Search

## Vendors and cost where applicable

Searcher: SCOTT, L. C. 4939 NA Sequence (#) \_\_\_\_\_ STN \_\_\_\_\_  
 Searcher Phone #: \_\_\_\_\_ AA Sequence (#) \_\_\_\_\_ Dialog \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_ Structure (#) \_\_\_\_\_ Questel/Orbit \_\_\_\_\_  
 Date Searcher Picked Up: \_\_\_\_\_ Bibliographic \_\_\_\_\_ Dr. Link \_\_\_\_\_  
 Date Completed: 04-20-00 Litigation \_\_\_\_\_ Lexis/Nexis \_\_\_\_\_  
 Searcher Prep & Review Time: 25 Fulltext: \_\_\_\_\_ Sequence Systems \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_ Patent Family \_\_\_\_\_ WWW/Internet \_\_\_\_\_  
 Online Time: 20 Other \_\_\_\_\_ Other (specify) CSA

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2000, 01:03:34 ; Search time 417.26 Seconds  
(without alignments)  
-5734.244 Million cell updates/sec

Title: US-09-203-548-2  
Perfect score: 788  
Sequence: 1 GCCGCCGAACCGCGCGCC.....AATATTAGAAAGTTTGAGC 788

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 821193 seqs, -1518192014 residues  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_v1:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_v1:\*  
32: gb\_hcg1:\*  
33: gb\_hcg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_hcg3:\*  
42: gb\_hcg4:\*  
43: gb\_hcg5:\*  
44: gb\_hcg6:\*

45: gb\_hcg7:\*  
46: em\_hcg1:\*  
47: em\_hcg2:\*  
48: em\_hcg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	679.8	86.3	1941	9	HSPROGBIN	Y12711 H.sapiens m
2	563	71.4	1893	3	SSSTERMBP	X99714 S.scrofa MR
3	542	68.8	1786	12	AF042491	AF042491 Mus muscu
4	519.6	65.9	678	12	RNO5837	AJ005837 Rattus no
5	508.6	64.5	1885	12	RNU63315	U63315 Rattus norv
6	411	52.2	121803	40	AC004835	AC004835 Homo sapi
7	200.4	25.4	1874	9	HSBJ2030	AJ002030 Homo sapi
8	198	25.1	180439	41	AC011193	AC011193 Homo sapi
9	71.8	9.1	930	8	AF153283	AF153283 Arabidops
10	60.2	7.6	847	8	AF153284	AF153284 Arabidops
11	55.2	7.0	2209	7	SCPL170W	27326 S.cerevisia
12	55.2	7.0	55786	7	SCLACHXVI	X96770 S.cerevisia
13	54.8	7.0	103495	8	ATAC006585	AC006585 Arabidops
14	52.8	6.7	110070	41	AC009340	AC009340 Drosophill
15	52.2	6.6	34851	35	CELK07E3	U41552 Caenorhabdi
16	51.8	6.6	39616	8	SPAC2588	AL133225 S.pombe c
17	51.8	6.6	41100	8	SPAC26H5	Z99126 S.pombe chr
18	51.4	6.5	724	40	AF173937	AF173937 Homo sapi
19	51.2	6.5	82812	8	ATT21J18	AL132963 Arabidops
20	48	6.1	2243	1	SERERYFGH	M54983 S.erythraea
21	48	6.1	4254	1	SERERVABIO	X60379 S.erythraea
22	47.8	6.1	10035	2	AE001999	AE001999 Deinococc
23	47.8	6.1	38209	2	SC5G9	AL117385 Streptomy
24	47.2	6.0	37730	1	SC59	AL049841 Streptomy
25	46	5.8	2156	1	SLLMRAG	X59926 S.lincoln
26	46	5.8	26440	1	SCE15	AL049707 Streptomy
27	46	5.8	36270	1	SLLINC	X79146 S.lincoln
28	45.4	5.8	37200	1	SCE63	AL035640 Streptomy
29	45.4	5.8	190014	7	AP000836	AP000836 Oryza sat
30	45.4	5.8	325160	43	AC015817	AC015817 Homo sapi
31	45.2	5.7	40344	1	SCH5	AL035636 Streptomy
32	45	5.7	28795	1	AVINIFC	M20568 A.vinelandi
33	44.8	5.7	504	11	AF049872	AF049872 Homo sapi
34	44.6	5.7	40024	1	SC5F7	AL096872 Streptomy
35	44.2	5.6	2951	2	AF123319	AF123319 Streptomy
36	44.2	5.6	43147	1	SC4A10	AL109663 Streptomy
37	44.2	5.6	52149	43	AC012958	AC012958 Drosophill
38	44.2	5.6	139825	44	AC010919	AC010919 Drosophill
39	44.2	5.6	159606	44	AC008336	AC008336 Drosophill
40	44	5.6	4324	2	AF076477	AF076477 Burkholde
41	43.8	5.6	493	1	SGU72183	U72183 Streptomyce
42	43.8	5.6	5251	14	AF000429	AF000429 Shuttle c
43	43.8	5.6	11143	1	SPU23762	U23762 Streptomyce
44	43.8	5.6	40105	1	SC5F2A	AL049587 Streptomy
45	43.8	5.6	139825	43	AC010919	AC010919 Drosophill

ALIGNMENTS

RESULT 1  
HSPROGBIN HSPROGBIN 1941 bp mRNA PRI 07-AUG-1998  
LOCUS H.sapiens mRNA for putative progesterone binding protein.  
DEFINITION Y12711  
ACCESSION Y12711 GI:2062021  
VERSION Y12711.1  
KEYWORDS progesterone binding protein.



QY	72	GC	GGGGCCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGTTGGTGGCGACTGGCGC	131
DB	14	GAG	CCCTCCCAACCTTTGCTCCAGCATCATGGCTGCCGAGGATGTGGCGCTACCGGGC	73
QY	132	CG	ACCCAAAGCGATCTGGAGAGCGGGGCTGCTGCATGAGATTTTCACGTGCCCGCTCAA	191
DB	74	CG	ACCCGAGCGAGCTAGAGGGCGGGGCTGCTGCATGAGATTTTCACGTGCCCGCTCAA	133
QY	192	CT	GTGCTGCTCTTTGGCCTCTGCAATCTTCCTGCTCTACAAGATCGTGGCGGGGACACGCG	251
DB	134	CT	GTGCTGCTCTGGGCTCTGCAATCTTCCTGCTCTACAAGATCGTGGCGGGGACACGCG	193
QY	252	GG	CGGGCAGCGGCGACGAGCAGCAGCGANGCGCCCCCTCTGCCCCGCGCTCAAGCGGCG	311
DB	194	GG	CGGGCAGCGGATAGC--GACGACGACGA--GCGCGCCCGCTGCCCGCTTAAGCGGGC	250
QY	312	CG	ACTTACCCCCCGCGAGCTCGGGCGCTTCGACGCGGCTCCAGGACCCCGCGCATACTCAT	371
DB	251	CG	ACTTACCCCCCTGCGAGAGTGGCTTCGACGCGGCTCCAGGACCCCGCTATACTCAT	310
QY	372	GG	CATCAACGCCAAGGTGTTCGATGTGACAAAGCGCCAAATTCACGGGCCCGAGGG	431
DB	311	GG	CATCAACGCCAAGGTGTTCGACGTGACCAAGCGCGCAAGTTCACGGGCCCGAGGG	370
QY	432	GC	CGTATGGGGTCTTTGCTGGAAAGAGATGCATCCAGGGGCGCTTGCCACATTTTGCTGGA	491
DB	371	GC	GTACGGGCTTTGCTGGAAAGAGACGATCCAGGGGCGCTGGCCACGTTTGCTGGA	430
QY	492	TA	AGGAACACTAGAGATGAGTACGATGACCTTCTGCACCTCAGTGTGCCCGACGAGA	551
DB	431	TA	AGGAAGCCCTGAAGACAGGATGATGACCTTCTGCACCTCAGTGTGCCCGACGAGA	490
QY	552	GACT	CTGAGTGACTGGGAGTCTCAGTTCACTTTCAAGTATCATCAGTGGGCAAACTGCT	611
DB	491	GACC	TGAATGACTGGGACTCTCAGTTCAGTTCAGATACCATCAGTGGGCAAACTGCT	550
QY	612	GA	AGGAGGGGAGAGCCCCACTGTACTCAGATGAGGAAGAACCAAAAGATGAGAGTTC	671
DB	551	GA	AGGAGGGGAGAGAGCCCCACTGTACTCAGATGAGGAAGAGCCCAAGATGAGAGCG	610
QY	672	CC	GGAAAAATG-TTAAAGCATTCAGTGGAGATATATCTA-TNNTGATTTTGCAAAAATCA	729
DB	611	TC	GGAAAAATGATTAAGCGCTTCGGTGGAGCATCTATTTTGTATTTTCGAAATCA	670
QY	730	TT	GTAACTGCCACTGTGCTTTTAAACATAGTG-TTCAATATTTGAAAGTTT	784
DB	671	TT	GTAACTGCCACTGTGCTCTTAAACATGTTGATTTCAATATTTAGAGAGTT	726
RESULT	3			
LOCUS	AF042491			
DEFINITION	AF042491	1786 bp	mus musculus putative membrane associated progesterone receptor component mRNA, complete cds.	22-JAN-1998
ACCESSION	AF042491			
VERSION	AF042491.1			
KEYWORDS	GI:2801792			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 1786)			
TITLE	Wong, S., Lunn, R.M., O'Brien, D.A., Bell, D.A. and Eddy, E.M.			
JOURNAL	The expression of a putative membrane associated progesterone receptor component in the mouse testis and epididymis			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 1786)			
TITLE	Wong, S., Lunn, R.M., O'Brien, D.A., Bell, D.A. and Eddy, E.M.			
JOURNAL	Direct Submission			
FEATURES	Submitted (12-JAN-1998) LRDT, NIEHS, NIH, 111 Alexander Drive, C4-04, Research Triangle Park, NC 27709, USA			



[illegible]

```

RESULT      4
RNO5837      678 bp      mRNA      ROD      08-MAY-1998
LOCUS      Rattus norvegicus mRNA for putative progesterone binding protein.
DEFINITION  AJ005837
ACCESSION  AJ005837.1 GI:3127856
VERSION    progesterone binding protein; putative.
KEYWORDS   Norway rat.
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
            Noelte.I.
            Direct Submission
            Submitted (28-APR-1998) Noelte I., Biochemiezentrum Heidelberg, Inf
            328, 69120 Heidelberg, GERMANY
            2 (bases 1 to 678)
REFERENCE   Noelte.I., Sohn,K., Wegehingl,S. and Wieland,F.
AUTHORS    Rat homologue to a putative progesterone binding protein :
            molecular characterization and localization
TITLE      Unpublished
JOURNAL    Location/Qualifiers
FEATURES   1..678
            /organism="Rattus norvegicus"
            /strain="fisher 344"
            /db_xref="taxon:10116"
            /tissue_type="liver"
            75..662
            /gene="Lewi"
            75..662
            /gene="Lewi"
            /codon_start=1
            /product="putative progesterone binding protein"
            /protein_id="CAA06732.1"
            /db_xref="GI:3127857"
            /db_xref="GI:3127857"
            /translation="MAEDVAVTADGPSELEGGGLQEIFTSPLNLLGLCFLVLY
            IVRGDQPGASGDNDDPEPLPLKPRDPTPALRRYDGVQDPRILMAINGKVDYTK
            GRKFYGPFGYVAGRADSRGLATCFLDKEALKDEYDDLSTLPAQQTFLNDWDSQF
            TKYIHVGKLLKEGEPTVYSDDEEPKDEARKSD"
            misc_signal 270..287
            /gene="Lewi"
            /function="potential nuclear insertion signal"
            630..653
            /gene="Lewi"
            /function="potential nuclear insertion signal"
            151 a 192 c 205 g 130 t
            BASE COUNT 151 a 192 c 205 g 130 t
            ORIGIN
            Query Match 65.9%; Score 519.6; DB 12; Length 678;
            Best Local Similarity 87.8%; Pred. No. 1.3e-94;
            Matches 567; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
            QY 36 GGAGGAGAAAGTGGCGAGTCCGGATCCCTGCTAGCGGGCCCAACCTTTACTCCAGA 95
            DB 11 GGCACAGAGGCGGACTGTTCGGATCTCTGCATACAGGCGCCCAACCTTTGCTCCAGA 70
            QY 96 GATCATGGCTGCCGAGGATGTGGTGGCGACTGGCGCGCACCCCAACGATCTGGAGAGCGG 155
            DB 71 GATCATGGCTGCCGAGGATGTGGTGGCGACTGGCGCGCACCCCAACGATCTGGAGGCGG 130
            QY 156 CGGGCTGCTGCATGAGATTTTACGTGCGCGCTCAACCTGCTGCTGCTGCGCTCTGAT 215
            DB 131 CGGGCTGCTTCAAGAGATTTTACGTGCGCTCTCAACCTGCTGCTGCTGCGCTCTGAT 190
            QY 216 CTTTCCTGCTCTACAAGATCGTGGCGGGGACCAAGCGCGCGCCAGCGACGAC 275
            DB 191 CTTTCCTGCTCTACAAGATCGTTCGGGGGACCAAGCGCGGTGCTTCAACGCGGACGACGAC 250
            QY 276 GACGANGCCGCCCTCTGCGCCGCCCTCAAGCGGCGCGACTTCAACCCCGCGGAGCTGGC 335

```

Query Match		64.5%	Score 508.6;	DB 12;	Length 1885;
Best Local Similarity		82.3%	Pred. No. 2.1e-92;		
Matches 631; Conservative		0;	Mismatches 128;	Indels	8; Gaps
QY	25	GCTCGCTCAGAGGAGGAGAAAGTGGCGAGTTCCTGGATCCCTGCTAGCGCGGCCCAACC	84		
DB	24	GTTTCGCTTAGAGGCGGAGGAAGCCGACTGTTTCGG-TCTCTCATAAAGGCCCAACC	82		
QY	85	TTTACTCCACAGATCATGGCTGCGGAGGATGTGTGGCGACTGGCGCCGACCCCAACGCGAT	144		
DB	83	TTTGTCTCCAGAGATCATGGCTGCGGAGGATGTGTGGCGACTGGCGCCGACCCACGCGAG	142		
QY	145	CTGGAGAGCGCGGGCTGCTGCATGAGATTTTCACGTTCGCGCTCAACCTGTGCTGCTT	204		
DB	143	CTGGAGGCGCGGGCTGCTTCAAGAGATTTTCACGTTCGCTCTCAACCTGTGCTGCTT	202		
QY	205	GGCTCTGCATCTTCTGCTCTACAAGATCGTGGGGGACACAGCGGGGGCCAGCGGC	264		
DB	203	GGCTCTGCATCTTCTGCTCTACAAGATCGTGGGGGACACAGCGGGGGCCAGCGGC	262		
QY	265	GACAGACGACGACGAGCGCGCCCTCTGCCCGGCTCAAGCGCGCGACATTCACCC	324		
DB	263	GACAGACGACGACGAGCGCGCCCTCTGCCCGGCTCAAGCGCGGTACATTCACCC	322		
QY	325	GCCGAGCTGGGGGCTTCGACGGCTCCAGAGCCCGCGCATCTCATGCCATCAACGCG	384		
DB	323	GCCGAATTAAGCGATACGATGGAGTCCAGGACCCGCGCATTTCTATGCCATCAACGCG	382		
QY	385	AAGGTGTTTCGATGTGACCAAGCGCGAAATTCACGGGCGCGAGGGGCGGTATGGGGTC	444		
DB	383	AAGGTGTTTCGATGTGACCAAGCGCGAAATTCACGGGCGCGAGGGGCGGTATGGGGTC	442		
QY	445	TTTGTCTGAGAGATGCATCCAGGGGCTTGCCACATTTTGCCTGATGAAGNAGCACTG	504		
DB	443	TTTGTCTGAGAGATGCATCCAGGGGCTTGCCACATTTTGCCTGATGAAGNAGCACTG	502		
QY	505	AAGGATGAGTACGATGACCTTTCTGACCTCACCTGCTGCCAGCAGGAGACTCTGAGTGAC	564		
DB	503	AAGGATGAGTATGATGACCTTTCTGACCTCACCTGCTGCCAGCAGGAGACTCTGAGTGAC	562		
QY	565	TGGAGTCTCAGTTTCAAGTATCATCACGT-TGGGAACTGCTGAAGGAG	618		
DB	563	TGGAGTCTCAGTTTCAAGTATCATCACGT-TGGGAACTGCTGAAGGAG	622		
QY	619	GCGGAGGAGCCACTGTGACTCAGATGAGGAGACCAAAAGATCAGAGTTCCCGGAA	678		
DB	623	GCGGAGGAGCCACTGTGACTCAGATGAGGAGACCAAAAGATCAGAGTTCCCGGAA	682		
QY	679	AATGTTAAAGCATTCAGTGAAGTATATCTATNTGTATTTTGCAAAATCATTTGTAACA	738		
DB	683	GTGACTGAAGCATTCAGTGAAGTATATCTATNTGTATTTTGCAAAATCATTTGTAACA	742		
QY	739	GTCCACTNTGCTTTAAACATAGTG-TTACAAATATTAGAAAGTTT	784		
DB	743	TTCCAGTCTCTTAAACATAGTG-TTACAAATATTAGAAAGTTT	789		
RESULT	6				
AC004835					
LOCUS	AC004835				
DEFINITION	Homo sapiens clone DJ0555N02, complete sequence.			PRI	24-MAR-1999
ACCESSION	AC004835				
VERSION	AC004835.2	GI:4508154			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 121803)				
	Waterston, R.H.				
	The sequence of Homo sapiens clone				

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 121803)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 121803)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (24-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Mar 24, 1999 this sequence version replaced gi:3213162.
FEATURES	Location/Qualifiers
source	1..121803
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="DU055N02"
BASE COUNT	35035 a 24906 c 25571 g 36291 t
ORIGIN	
Query Match	52.2%; Score 411; DB 40; Length 121803;
Best Local Similarity	91.4%; Pred. No. 7.9e-73;
Matches 495; Conservative	0; Mismatches 41; Indels 0; Gaps 0;
Qy 1	GCSCCGCAACCCGCGCGCCACCTCGCTCGCTCAGAGGAGGAGAAAGTGGCGAGTTCCGG 60
Db 33823	CGCCCGCAACCCGCGCGCCACCTCGCTCGCTCAGAGGAGGAGAAAGTGGCGAGTTCCGG 33882
Qy 61	ATCCCTGCCTAGCGCGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGTGGTG 120
Db 33883	ATCCCTGCCTAGCGCGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGTGGTG 33942
Qy 121	CGGACTGGCGCGCACCAACGATCTCGAGAGCGCGGGCTGTCATGAGATTTTCACG 180
Db 33943	CGGACTGGCGCGCACCAACGATCTCGAGAGCGCGGGCTGTCATGAGATTTTCACG 34002
Qy 181	TCGCCGCTCAACCTGCTGCTGTGGCCTTCGATCTTCCTGCTCTACAGATCGTGGC 240
Db 34003	TCGCCGCTCAACCTGCTGCTGTGGCCTTCGATCTTCCTGCTCTACAGATCGTGGC 34062
Qy 241	GGGACACAGCGCGCGCGCAGCGGACAGACGACGACGCGCCCTCTGCCCGCG 300
Db 34063	GGGACACAGCGCGCGCGCAGCGGACAGACGACGACGCGCCCTCTGCCCGCG 34122
Qy 301	CTCAAGCGCGCGGACTTCAACCCCGCGCGAGCTCGCGCGCTTCGAGCGCGTCCAGGACCG 360
Db 34123	CTCAAGCGCGCGGACTTCAACCCCGCGCGAGCTCGCGCGCTTCGAGCGCGTCCAGGACCG 34182
Qy 361	CGCATCTCATGGCCATCAACGGCAAGGTGTTCGATGTACCAAGGCGCGCAATTTCTAC 420
Db 34183	CGCATCTCATGGCCATCAACGGCAAGGTGTTCGATGTACCAAGGCGCGCAATTTCTAC 34242
Qy 421	GGGCGCGAGGGCGGTATGGGCTTTTCTGGAGAGATGCATCCAGGGCGCTTGC 476
Db 34243	GGGCGCGGTACGCGCGCGGCGGCGGCGCTTGGAGACAAAGAGGGGGCCCCCGC 34298
RESULT 7	
HSJ2030	HSJ2030 1874 bp mRNA PRI 30-DEC-1998
LOCUS	Homo sapiens mRNA for putative progesterone binding protein.
DEFINITION	Homo sapiens mRNA for putative progesterone binding protein.
ACCESSION	AJ002030
VERSION	AJ002030.1 GI:2570006
KEYWORDS	progesterone binding protein.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.
	1 (bases 1 to 1874)
	Gerdts,D.



[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

1 (bases 1 to 847)  
Choi, J. H., Choi, H. and Gray, P.  
Plant homologues of mammalian putative progesterone-binding  
membrane proteins  
Unpublished  
2 (bases 1 to 847)  
Choi, J. H., Choi, H. and Gray, P.  
Direct Submission  
Submitted (19-MAY-1999) School of Biology, Georgia Institute  
Technology, MC 0230, Atlanta, GA 30332-0230, USA  
Location/Qualifiers

JOURNAL Submitted (19-MAY-1999) School of Biology, Georgia Institute of Technology, MC 0230, Atlanta, GA 30332-0230, USA

FEATURES

Location/Qualifiers

1..847

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/db\_xref="dbEST:N37902"

/chromosome="5"

1..847

/gene="Atmpl"

62..724

/gene="Atmpl"

/codon\_start=1

/product="putative progesterone-binding protein homolog"

/protein\_id="AAD34616.1"

/db\_xref="GI:4960156"

gene

CDS



/db\_xref="taxon:4932"  
/chromosome="XVI"  
/map="left arm"  
/clone\_lib="pHC79 or Supercos derivative cosmids (L. Riley  
St Louis, Missouri)"  
/clone="cos9574"  
/clone="cos8484"  
complement(1..542)  
/gene="OYE3"

gene

/db\_xref="SGD:S0006092"  
complement(<1..542)  
/gene="OYE3"

CDS

/note="129279"

/codon\_start=1

/db\_xref="SGD:S0006092"

/product="P2291 protein"

/protein\_id="CAA65573.1"

/db\_xref="GI:1403538"

/db\_xref="SWISS-PROT:P41816"

/translation="MPFVKGFPISLRDNLPEPIKIGTQLAHRAVMPPLTMRATH  
PCNIPKHAAYVYQORAPGTMITEGTFISQAGGDNAPGINSQVQAEWKNIF  
LAHQCCQFAWQSLGWSAFPDLARDGLRYDCASDRVYMNATLQERAKADANNLEH  
SITKDDIKQYIKDYTHAANS"  
1485..1943

CDS

/note="putative"

/codon\_start=1

/product="P2515 protein"

/protein\_id="CAA65551.1"

/db\_xref="GI:1403539"

/db\_xref="SPTREMBL:Q12091"

/translation="MSFIKNLLFGVKTSEDPTGLTGNGASNTDNDKSGEPVAVGNE  
FPTLSKNGHDEKIFAIIRGVYDCTRGROFYGSPGYTTFNAGHDHAGLNSFD  
LDVINDQDIPDLDTKEQDADLEWQHEFNKYPCIGTLIPEGVN"  
complement(2210..4009)

CDS

/note="putative"

/codon\_start=1

/product="P2520 protein"

/protein\_id="CAA65552.1"

/db\_xref="GI:1403540"

/db\_xref="SPTREMBL:Q99257"

/translation="MSGFHNVGNNMMAQQOQNRKIKISVRNQWATMNDLINFISR  
NARVAYDAHVGPVIGVYVNSGAEAESLMKNGVRFAGSNLKFELLDNNGASAGTSD  
TISLGRVLLKRYDPOTKLLNLGALHSDPELLOKGVFSSITQSKMPPAMKLAETK  
SLIVSVNLADNQLKDISAISILAQTFFPNLKNLANNQIFRFSLEVWKNKFKDURE  
LMTNNPIITDKLYTLEMLRFPKLVLDNVIVRDEQKLTQTVISLPMKIQOFEFFENDA  
LQOSTDFATNLNMDNNREQLNLYSPQSOFVSVDSTIPPSTVTDSDQPAFGYY  
MSSRNISKSVSSEKSTQOQLSISQOESINSIFKTLPTKHHLOBPNEYSWETISYQI  
NGFVITLHGFEETGKPELESNKTGKNYQKNRYNGYNSYNNKLSKKSFDRTWV  
IYPMNSVLIASDLLTVRAYSGAWKTASIAAQPPQQAASVLPQVASHNPNITTPQ  
POPSVYVPGNSIPGAPQGAWVWAPTQLQPPDVQSRNLNPVOLELLNKLHLETKLNAEY  
FMLAFQSNMYEVAIKGFQSMNGIPREAFVQ"  
4388..5680

CDS

/note="putative"

/codon\_start=1

/product="P2525 protein"

/protein\_id="CAA65553.1"

/db\_xref="GI:1403541"

/db\_xref="SPTREMBL:Q12467"

/translation="MTVLVTSASLKKMKCLAFNMGMNVCVTVSHASGGAKTGGRNVF  
NIFDSKTPSVRIKAFKNTIYOSAMGKGTKFSAMEINLITSLVRGYKGEKKNAINP  
LQTNVQILNKLHLETKLHLDKIDLEGNNLAAGPVNVAIPDITPQEEKKVKVLELNRAE  
NNDLPSRKKHLEKLLSLNDLMDCNDEEYQKISLYLQKNEESRTSVGASQQNHVID  
INSLARKQIKARQKSAIDQKKNQARIYQWNTQSFSEFSLVSGNILLFKREPNI  
LWKRLQNGSVFLGSGGKSKTKTKVLOGNLILHLSLENNKMDTLNNDHDSVFN  
NFTDLGVINASGSPDRVLEINEIELGKMKVCGNLYDNKNKIVVQSNPLLEDIKI  
PKSSTNSRFLISALLASFAFYRILSRQESKK"

gene

complement(5764..10278)

/gene="REV3 (also called PS01)"

/db\_xref="REV3 (also called PS01)"

/db\_xref="SGD:S0006088"

complement(5764..10278)

/gene="REV3 (also called PS01)"

/note="M29683"

/codon\_start=1

CDS

/db\_xref="SGD:S0006088"  
/product="P2535 protein"  
/protein\_id="CAA65554.1"  
/db\_xref="GI:1403542"

/db\_xref="SWISS-PROT:P14284"

/translation="MSRESNDTIQSDTVSRSSKSDYFRIQLNQDYMYSKPTFLDPSSH  
GESLPLNQFSQVNIIRVFGALPTGHVCHVGLPYMFYKYGQITDTSLRHORCA  
OVHKTLEVIRASFEBKBDKDLADGLGNLNFADVSVVKGIPYGVHVGNNLEYK  
ISLNFSCLSRISSELIRDGKIFGKKEIYESHIPYLLQWADENLFGCSWINDRCYF  
RSPVLNLIIDIKLTIINDQLLDLDFCDFKNVLSRDPFRVNGNGLIEDILDPQFIK  
NREKLQRDIHHDLEKLGDISIDIPKYPVSSARDMINELTMRQELSKEYKEPPT  
KRHSQHQSGEFAFYKKAQKHTSTFDGQIPNFENFDKMKQSAINTKRSYSWLP  
LWPLRQIETNNNSMODKNDQVNASFEYEICQVDNENEGVKGNSIKRSYSWLP  
STASPKDSIILLDHQTKYHNTINFSDCANTQMAKSKRLRSSVSANKTSLSRKRRK  
VMAAGLRIGKRAFVYGEPPGYQDILNKLEDEGFPKIDYKDFPSPVDQENPYAYA  
GKFEISSTHVSTRIPVFGGETVSVYNKPTDFMSWKYALKPPTVDAQWYKNVP  
SMGNKTESQISMHTPHSKFTFLPDADVGKQKRSVSHDSLTLTLEHANTRSK  
IPDAIDEMVSIWCLEETFLPDLDIAYGIMVHKASDSTFPFKIQHCINEIPVM  
FYSEFEMTEALDVLVLDLDPDILSGFEIHNSGVIIERCQKHOFDIYRELARVC  
OIKTSLDSTGYAHSSGIMTGRHMINIRALSDVNLTOYTIESAFTLHKRLPHF  
SPESLNMNNAKSTTELKTVLNWLSRAQINQLLKKQDIARNTIEQALIGIDFHS  
VYTRGQFKVSEFLIRICKSEFILLSPGKKDVKQKALECVPLVMEPESAFYKNDVTIA  
VLDQSLPSIMIGYCYSTMIGRVREINLTENNIGVSKFSLPRNTLALLKNDVTIA  
PNGVYAKTSVRKSTLSMLTDILDVRVMIKKTMEIGDDNTLKRLLNKLALKLL  
ANVTGYTSAFSRMPCLADSIQVGTRETLEKADIIEKDETWNKAVYVGDTSL  
FYVLPKTAIEAESIGHAARVTONNPRFIFLKEKYHPSILI SKKRVGSGYESP  
SQPLPFDAGIETVRDGPQAQKIIERCIIRLLFTQDKLSIKKYLQNEFFKIQIGK  
VSAQDFCAKEVGLKAYKSEKAPAGAVVYKRNEDHRAEQYKERIPYLVVKGQK  
QLLRERCVPPEELEGLENDSEYINKILIPDLRFLNLGINVGNWAQAEIVKSR  
ASPTTKVENVITRVGTSATCCNGEELTKICSLQDCDCKLEKSTTTLFLINKLRQ  
KEYQTLKTVCTCSYRYTSDAGIENDHIAKSCNSYDCPVYSRVAERYLNDQSVQR  
EALISLNDN"

CDS

10309..11150

/note="putative"

/codon\_start=1

/product="P2540 protein"

/protein\_id="CAA65555.1"

/db\_xref="GI:1403543"

/db\_xref="SPTREMBL:Q12092"

/translation="MIMNSTVVIYIKVGRKRPQGLDPPKFEWNGTKERQLWTMVSN  
LNE SQDILQWNLKIFTEPEFFLKRTYKLFABHLELOLEKHLQLEKHLQTSNDQVN  
EGMSDLHKYPTLQNDNLLNVSAPLTERODSEEVETVENAOLHLOTSKLINH  
KKTSDENPNKLDKDGINKENECESSDDDLSSLSVSKSALEALMDRLQF"

CDS

complement(11127..12248)

/note="putative"

/codon\_start=1

/product="P2545 protein"

/protein\_id="CAA65556.1"

/db\_xref="GI:1403544"

/db\_xref="SPTREMBL:Q12529"

/translation="MTIDGDVHEISPPFVQVROTQWGRACFSNGNIPKGTVLQVSNF  
TGTISIVERPRKEVCHNCFAYANAKTKYKLYNDYLDRLVCNAHQINPKFLGALWF  
CSEHCRTSYLIQINIIETELCEYILLHHFESMLKRYNTSEQEKNLSILSENVTQS  
SWDEIESKWIPRINNMKSAKRNQLPPTCEDEYCCIRFCESLFLNKLMPDQCITYRA  
FNNLQSNELSKISKFPVLLHFQKLVFTYLLPPLSHLRMLSLPLRLHGTETYNFA  
GLWQEGEASDSREYFGYWFPEASFYFNHSCNPNTIKYKGNMLFTMNRDIKKDEQIC  
IDYSGLDLPTVKRAFLADSWFFDCACERCKSELOSVH"

complement(12521..14668)

/note="putative"

/codon\_start=1

/product="P2550 protein"

/protein\_id="CAA65557.1"

/db\_xref="GI:1403545"

/db\_xref="SPTREMBL:Q12083"

/translation="MSQHINKLSDSVSERLKSQACTVSLASAVREIVQNSVDAAHTTI  
DVNIDLPNSFAVYDDGIGLTRSDNLNLTQNTYSKIRKMDLVTKTYRGDALYS  
ISNVSLFKSKKKDYNASAMRRFSPKSVMLSENTILPIDFPWKICPWSRTKSGTVI  
VEDMLNLPVRRILKEEPPKFTNTIKADMLQILVMHPMISLNVQYDKLRINTYVL  
FRSKNITEGLTKHQOQSVLRNVFGALIPDMLKVLKENEYQIEGILSKMPVGLKD  
LQFIYINGRIADSAFCQGVDSLFQAQDFGEKGSMLKTKSVGKPRSHRPVFIIDVRC  
PQIIDDLLQDPARKIVKPSHIRTIEPLIVKTIKSLFTQGYLTLPDQKSDSSFEIVNC  
KTATLPDSRIQISKRQVLSNMKRIARINSYIGKPAVNGCRINNSINYEKINRID  
GQSKRLRNKLSRPYDSGTFEDYDSIGTKITITDFISIRSVLAKYEVINQVDKFTLIR

CDS

```

LQSIHNCPLLVLDQHCADIRLEELFYSLLEPVVTFVARDLKDCCEIVDRTEA
DLFKHYQSEKPKDGIYETIEGTMETSLLEIKTPEMLTSKYNGDKDYKLVKLQHAH
DXDFKTLKPKLGHFENYTSVDKLYWKKYSQCVPTVHEILNSKACRSAYMFGDELIR
QECILLISKLSRCHNPECAHGRPSMVPALX"
complement(15089..15871)
/gene="SVS1"
/db_xref="SGD:S0006084"
complement(15089..15871)
/gene="SVS1"
/note="D50278"
/codon_start=1
/product="P2554 protein"
/protein_id="CAA65558.1"
/db_xref="GI:1403546"
/db_xref="SWISS-PROT:Q12254"
/translation="MIFKILCSLLVTSNFSALYVNETSYTPYTKLTPYTSVSPQ
ETLTYSDETTTYTFTSYWFTTSQAALISTPASTPTASTPTASTPTSLTSTNEYTT
LYSDTDYTSITLSTYITLSTESANEKAEQISTVTEASTVTESGSYTSTLST
LLVTVNSQASNTIATAGDAASNDALKLVSAEHQSOMIQTTSADQYCSASTKY
VTVTAATVEVTTTAPVVKYVTTITADASNVTGSANNGHI"
complement(16377..17198)
/note="putative"
/codon_start=1

Query Match      7.08;  Score 55.2;  DB 7;  Length 55786;
Best Local Similarity 52.48;  Pred. No. 0.1; 128;  Indels 9;  Gaps 1;
Matches 151;  Conservative 0;  Mismatches 128;

QY 340 TTGCAGCGGTCAGGACCCGCGCATCTATCGCCATCAACGGCAAGGTGTTCGATGTG 399
Db 1638 TTTAACGGCCACGACGATGAATAATTTATTGCTATTAGGGGCAAGATACGACTGC 1697
QY 400 ACCAAGGCCCAATCTACGGCCGCGGCGGTATGGGCTTTTCGTGGAAGAGAT 459
Db 1698 ACAAGAGGGAGGAGCTTTACGGTCCAGCGGCCATACACTTAATTTCGAGGCCATGAT 1757
QY 460 GCATCCAGGGCCCTTGCCACATTTTGCTGATAGGAGCACTGAAGGATCAGTA--- 516
Db 1758 CGTCGGGTGCTTGCAATGTAACCTTCGATCGGACGTTATTAAGATGGGATCAG 1817
QY 516 ----CGATGACCTTTCTGACCTCACTGCTGCCAGGAGGACACTCTGAGTGACGTGGAG 570
Db 1818 CCTATCGATCCCTTAGATGATCTGACAAAGACAGATTCAGCCTGATGAGTGCGCA 1877
QY 571 TCTCAGTTCACTTCAAGTATCATCACCTGGGCGGCAAACTGCTGAAGAG 618
Db 1878 GAGCATTTTGAGATAAGTACCCATGCTGTTGCTACTCTGATCCGGAG 1925

RESULT 13
ATAC006585/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence,
complete sequence.
ACCESSION AC006585
VERSION AC006585.6 GI:4559344
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi.
REFERENCE
1 (bases 1 to 103495)
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
Fraser,C.M. and Venter,J.C.
Unpublished
TITLE Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence
REFERENCE
2 (bases 1 to 103495)
AUTHORS Lin,X. and Kaul,S.

```

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT

Direct Submission  
Submitted (19-FEB-1999) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA, xlinetigr.org  
3 (bases 1 to 103495)  
Lin,X.

Direct Submission  
Submitted (04-APR-1999) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Apr 4, 1999 this sequence version replaced gi:4522000.  
Address all correspondence to:  
Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr.  
Rockville, MD 20850, USA  
e-mail: xlinetigr.org  
BAC clone F27C12 is from Arabidopsis chromosome II and is near the  
molecular marker m238.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.

Genes were identified by a combination of three methods: Gene  
prediction programs including GRAIL (available by anonymous ftp  
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
Washington), Genscan (Chris Burge,  
http://www.cbs.dtu.dk/netgene/cbsnetgene.html), and NetPlantGene  
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the  
complete sequence against a peptide database and the Arabidopsis  
EST database at TIGR (http://www.tigr.org/cdb/at/at.html).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by more  
than two gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,  
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are  
identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of  
genomic sequence that are not annotated as genes but have predicted  
exons by GRAIL are annotated as misc features.

FEATURES  
source  
1. 103495  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="II"  
/map="m238"  
/clone="F27C12"  
100..136  
/rpt\_family="AT\_rich"  
repeat\_region  
206..243  
/rpt\_family="AT\_rich"  
complement(245..344)  
misc\_feature  
/note="exon predicted by xgrail, quality  
excellent\_shadowexon"  
262..349  
/note="exon predicted by xgrail, quality excellent"  
complement(353..418)  
repeat\_region  
/rpt\_family="AT\_rich"  
mRNA  
join(1069..1197,1624..1742,1826..1889,2061..2169,  
2257..2410,2506..2755,2850..2926,3010..3062,3149..3203,  
3287..3608)  
/gene="F27C12.1"  
1069..3608  
/gene="F27C12.1"  
join(1626..1742,1826..1889,2061..2169,2257..2410,  
2506..2755,2850..2926,3010..3062,3149..3203,3287..3475)  
/gene="F27C12.1"  
/codon\_start=1  
/product="protein phosphatase 2C"  
/protein\_id="AAD23006.1"  
/db\_xref="GI:4559345"  
/translation="MGTYLSPKTEKLSGDKLRFGLSSMOGWRTMEDAHAAIL."





REFERENCE  
AUTHORS

1 (bases 1 to 110070)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacieb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Sequencing of *Drosophila melanogaster*  
Unpublished  
2 (bases 1 to 110070)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacieb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Aug 27, 1999 this sequence version replaced gi:5734690.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgpe@fruitfly.berkeley.edu](mailto:bdgpe@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 103 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 831: contig of 831 bp in length  
\* 832 911: gap of unknown length  
\* 912 1816: contig of 905 bp in length  
\* 1817 1896: gap of unknown length  
\* 1897 2535: contig of 639 bp in length  
\* 2536 2615: gap of unknown length  
\* 2616 3542: contig of 927 bp in length  
\* 3543 3622: gap of unknown length  
\* 3623 4606: contig of 984 bp in length  
\* 4607 4686: gap of unknown length  
\* 4687 5239: contig of 553 bp in length  
\* 5240 5319: gap of unknown length  
\* 5320 5819: contig of 500 bp in length  
\* 5820 5899: gap of unknown length  
\* 5900 6999: contig of 1100 bp in length  
\* 7000 7079: gap of unknown length  
\* 7080 7734: contig of 655 bp in length  
\* 7735 7814: gap of unknown length  
\* 7815 8642: contig of 828 bp in length  
\* 8643 8722: gap of unknown length  
\* 8723 9398: contig of 676 bp in length  
\* 9399 9478: gap of unknown length  
\* 9479 10194: contig of 716 bp in length  
\* 10195 10274: gap of unknown length  
\* 10275 11187: contig of 913 bp in length  
\* 11188 11267: gap of unknown length  
\* 11268 12292: contig of 1025 bp in length  
\* 12293 12372: gap of unknown length  
\* 12373 13092: contig of 720 bp in length  
\* 13093 13172: gap of unknown length  
\* 13173 13737: contig of 565 bp in length  
\* 13738 13817: gap of unknown length  
\* 13818 14740: contig of 923 bp in length  
\* 14741 14820: gap of unknown length  
\*  
\* 14821 15506: contig of 686 bp in length  
\* 15507 15586: gap of unknown length  
\* 15587 16145: contig of 559 bp in length  
\* 16146 16225: gap of unknown length  
\* 16226 16757: contig of 532 bp in length  
\* 16758 16837: gap of unknown length  
\* 16838 17617: contig of 780 bp in length  
\* 17618 17697: gap of unknown length  
\* 17698 18485: contig of 788 bp in length  
\* 18486 18565: gap of unknown length  
\* 18566 19107: contig of 542 bp in length  
\* 19108 19187: gap of unknown length  
\* 19188 20112: contig of 925 bp in length  
\* 20113 20192: gap of unknown length  
\* 20193 21946: contig of 1754 bp in length  
\* 21947 22026: gap of unknown length  
\* 22027 23219: contig of 1193 bp in length  
\* 23220 23299: gap of unknown length  
\* 23300 24700: contig of 1401 bp in length  
\* 24701 24780: gap of unknown length  
\* 24781 26054: contig of 1274 bp in length  
\* 26055 26134: gap of unknown length  
\* 26135 27537: contig of 1403 bp in length  
\* 27538 27617: gap of unknown length  
\* 27618 28880: contig of 1263 bp in length  
\* 28881 28960: gap of unknown length  
\* 28961 30009: contig of 1049 bp in length  
\* 30010 30089: gap of unknown length  
\* 30090 31018: contig of 929 bp in length  
\* 31019 31098: gap of unknown length  
\* 31099 32472: contig of 1374 bp in length  
\* 32473 32552: gap of unknown length  
\* 32553 34213: contig of 1661 bp in length  
\* 34214 34293: gap of unknown length  
\* 34294 35199: contig of 906 bp in length  
\* 35200 35279: gap of unknown length  
\* 35280 36595: contig of 1316 bp in length  
\* 36596 36675: gap of unknown length  
\* 36676 38058: contig of 1383 bp in length  
\* 38059 38138: gap of unknown length  
\* 38139 39267: contig of 1129 bp in length  
\* 39268 39347: gap of unknown length  
\* 39348 40454: contig of 1107 bp in length  
\* 40455 40534: gap of unknown length  
\* 40535 41619: contig of 1085 bp in length  
\* 41620 42972: contig of 1273 bp in length  
\* 42973 43052: gap of unknown length  
\* 43053 44592: contig of 1540 bp in length  
\* 44593 44672: gap of unknown length  
\* 44673 45891: contig of 1219 bp in length  
\* 45892 45971: gap of unknown length  
\* 45972 47685: contig of 1714 bp in length  
\* 47686 47765: gap of unknown length  
\* 47766 48974: contig of 1209 bp in length  
\* 48975 49054: gap of unknown length  
\* 49055 51039: contig of 1985 bp in length  
\* 51040 51119: gap of unknown length  
\* 51120 52367: contig of 1248 bp in length  
\* 52368 52447: gap of unknown length  
\* 52448 54651: contig of 2204 bp in length  
\* 54652 54731: gap of unknown length  
\* 54732 56987: contig of 2256 bp in length  
\* 56988 57067: gap of unknown length  
\* 57068 59377: contig of 2310 bp in length  
\* 59378 59457: gap of unknown length  
\* 59458 62048: contig of 2591 bp in length  
\* 62049 62128: gap of unknown length  
\* 62129 64732: contig of 2604 bp in length  
\* 64733 64812: gap of unknown length  
\* 64813 67277: contig of 2465 bp in length  
\* 67278 67357: gap of unknown length  
\* 67358 70450: contig of 3093 bp in length

us-09-203-548-2.rge		us-09-203-548-2.rge		Page 13	
70451	70530: gap of unknown length	70451	70530: gap of unknown length	70451	70530: gap of unknown length
70531	77373: contig of 6843 bp in length	70531	77373: contig of 6843 bp in length	70531	77373: contig of 6843 bp in length
77374	77453: gap of unknown length	77374	77453: gap of unknown length	77374	77453: gap of unknown length
77454	78134: contig of 681 bp in length	77454	78134: contig of 681 bp in length	77454	78134: contig of 681 bp in length
78135	78214: gap of unknown length	78135	78214: gap of unknown length	78135	78214: gap of unknown length
78215	78804: contig of 590 bp in length	78215	78804: contig of 590 bp in length	78215	78804: contig of 590 bp in length
78805	78884: gap of unknown length	78805	78884: gap of unknown length	78805	78884: gap of unknown length
78885	79553: contig of 669 bp in length	78885	79553: contig of 669 bp in length	78885	79553: contig of 669 bp in length
79554	79333: gap of unknown length	79554	79333: gap of unknown length	79554	79333: gap of unknown length
79634	80345: contig of 712 bp in length	79634	80345: contig of 712 bp in length	79634	80345: contig of 712 bp in length
80346	80425: gap of unknown length	80346	80425: gap of unknown length	80346	80425: gap of unknown length
80426	80975: contig of 550 bp in length	80426	80975: contig of 550 bp in length	80426	80975: contig of 550 bp in length
80976	81055: gap of unknown length	80976	81055: gap of unknown length	80976	81055: gap of unknown length
81056	81713: contig of 658 bp in length	81056	81713: contig of 658 bp in length	81056	81713: contig of 658 bp in length
81714	81793: gap of unknown length	81714	81793: gap of unknown length	81714	81793: gap of unknown length
81794	82406: contig of 613 bp in length	81794	82406: contig of 613 bp in length	81794	82406: contig of 613 bp in length
82407	82486: gap of unknown length	82407	82486: gap of unknown length	82407	82486: gap of unknown length
82487	83283: contig of 797 bp in length	82487	83283: contig of 797 bp in length	82487	83283: contig of 797 bp in length
83284	83363: gap of unknown length	83284	83363: gap of unknown length	83284	83363: gap of unknown length
83364	83964: contig of 601 bp in length	83364	83964: contig of 601 bp in length	83364	83964: contig of 601 bp in length
83965	84044: gap of unknown length	83965	84044: gap of unknown length	83965	84044: gap of unknown length
84045	84573: contig of 529 bp in length	84045	84573: contig of 529 bp in length	84045	84573: contig of 529 bp in length
84574	84653: gap of unknown length	84574	84653: gap of unknown length	84574	84653: gap of unknown length
84654	85192: contig of 539 bp in length	84654	85192: contig of 539 bp in length	84654	85192: contig of 539 bp in length
85193	85272: gap of unknown length	85193	85272: gap of unknown length	85193	85272: gap of unknown length
85273	85960: contig of 688 bp in length	85273	85960: contig of 688 bp in length	85273	85960: contig of 688 bp in length
85961	86040: gap of unknown length	85961	86040: gap of unknown length	85961	86040: gap of unknown length
86041	86615: contig of 575 bp in length	86041	86615: contig of 575 bp in length	86041	86615: contig of 575 bp in length
86616	86695: gap of unknown length	86616	86695: gap of unknown length	86616	86695: gap of unknown length
86696	87318: contig of 623 bp in length	86696	87318: contig of 623 bp in length	86696	87318: contig of 623 bp in length
87319	87398: gap of unknown length	87319	87398: gap of unknown length	87319	87398: gap of unknown length
87399	87933: contig of 535 bp in length	87399	87933: contig of 535 bp in length	87399	87933: contig of 535 bp in length
87934	88013: gap of unknown length	87934	88013: gap of unknown length	87934	88013: gap of unknown length
88014	88797: contig of 784 bp in length	88014	88797: contig of 784 bp in length	88014	88797: contig of 784 bp in length
88798	88877: gap of unknown length	88798	88877: gap of unknown length	88798	88877: gap of unknown length
88878	89531: contig of 654 bp in length	88878	89531: contig of 654 bp in length	88878	89531: contig of 654 bp in length
89532	89612: gap of unknown length	89532	89612: gap of unknown length	89532	89612: gap of unknown length
89612	90118: contig of 507 bp in length	89612	90118: contig of 507 bp in length	89612	90118: contig of 507 bp in length
90119	90198: gap of unknown length	90119	90198: gap of unknown length	90119	90198: gap of unknown length
90199	90875: contig of 677 bp in length	90199	90875: contig of 677 bp in length	90199	90875: contig of 677 bp in length
90876	90955: gap of unknown length	90876	90955: gap of unknown length	90876	90955: gap of unknown length
90956	91549: contig of 594 bp in length	90956	91549: contig of 594 bp in length	90956	91549: contig of 594 bp in length
Query Match		6.7%; Score 52.8; DB 41; Length 110070;			
Best Local Similarity		50.0%; Pred. No. 0.32;			
Matches 132; Conservative		0; Mismatches 132; Indels 0; Gaps 0;			
QY	354 GCACCGCGCATCTCATGCCATCAACGCGAAGGTTCGATGTGACCAAGGCCGCA 413	QY	354 GCACCGCGCATCTCATGCCATCAACGCGAAGGTTCGATGTGACCAAGGCCGCA 413	QY	354 GCACCGCGCATCTCATGCCATCAACGCGAAGGTTCGATGTGACCAAGGCCGCA 413
DB	71796 GACGGTCGCATCCTAGTGGCCATCTTTTCAACATCTAGCATGTCTCGCGATCCGGTCA 71855	DB	71796 GACGGTCGCATCCTAGTGGCCATCTTTTCAACATCTAGCATGTCTCGCGATCCGGTCA 71855	DB	71796 GACGGTCGCATCCTAGTGGCCATCTTTTCAACATCTAGCATGTCTCGCGATCCGGTCA 71855
QY	414 ATTCTACGGCGCGGCGCGGTATGGGTCTTTTGGTGAAGAGATGCATCCAGGGGCGCT 473	QY	414 ATTCTACGGCGCGGCGCGGTATGGGTCTTTTGGTGAAGAGATGCATCCAGGGGCGCT 473	QY	414 ATTCTACGGCGCGGCGCGGTATGGGTCTTTTGGTGAAGAGATGCATCCAGGGGCGCT 473
DB	71856 CTATTACGACGAATGGCGTTAATCCAACTATCGCGCGGATATTTCCAGGNAAT 71915	DB	71856 CTATTACGACGAATGGCGTTAATCCAACTATCGCGCGGATATTTCCAGGNAAT 71915	DB	71856 CTATTACGACGAATGGCGTTAATCCAACTATCGCGCGGATATTTCCAGGNAAT 71915
QY	474 TCCACATTTTGCCTGGATAAGGAAGACACTGAAGAGATGATGACCTTTCTGACCT 533	QY	474 TCCACATTTTGCCTGGATAAGGAAGACACTGAAGAGATGATGACCTTTCTGACCT 533	QY	474 TCCACATTTTGCCTGGATAAGGAAGACACTGAAGAGATGATGACCTTTCTGACCT 533
DB	71916 TTGATCAATTTCCCCCAAGGACTTTGAAGACAGCGAGGATTTTGATGATTGAGCAACCT 71975	DB	71916 TTGATCAATTTCCCCCAAGGACTTTGAAGACAGCGAGGATTTTGATGATTGAGCAACCT 71975	DB	71916 TTGATCAATTTCCCCCAAGGACTTTGAAGACAGCGAGGATTTTGATGATTGAGCAACCT 71975
QY	534 CACTGCTGCCAGCAGGACACTGTGAGTCTGGGAGTCTCAGTTCACTTCAAGTATCA 593	QY	534 CACTGCTGCCAGCAGGACACTGTGAGTCTGGGAGTCTCAGTTCACTTCAAGTATCA 593	QY	534 CACTGCTGCCAGCAGGACACTGTGAGTCTGGGAGTCTCAGTTCACTTCAAGTATCA 593
DB	71976 CTCACGCAACAAATGAACACATTACGAGATGGGACACGCGTACAAGATGAAGTACCC 72035	DB	71976 CTCACGCAACAAATGAACACATTACGAGATGGGACACGCGTACAAGATGAAGTACCC 72035	DB	71976 CTCACGCAACAAATGAACACATTACGAGATGGGACACGCGTACAAGATGAAGTACCC 72035
QY	594 TCACGTGGGCAAACTCTCTGAAGGA 617	QY	594 TCACGTGGGCAAACTCTCTGAAGGA 617	QY	594 TCACGTGGGCAAACTCTCTGAAGGA 617
DB	72036 ATTTGTTGAAGACTCAAGAGAA 72059	DB	72036 ATTTGTTGAAGACTCAAGAGAA 72059	DB	72036 ATTTGTTGAAGACTCAAGAGAA 72059
RESULT	15	RESULT	15	RESULT	15
LOCUS	CELK07E3/	LOCUS	CELK07E3	LOCUS	CELK07E3
DEFINITION	Caenorhabditis elegans cosmid K07E3.	DEFINITION	Caenorhabditis elegans cosmid K07E3.	DEFINITION	Caenorhabditis elegans cosmid K07E3.
ACCESSION	U41552	ACCESSION	U41552	ACCESSION	U41552
us-09-203-548-2.rge		us-09-203-548-2.rge		Page 13	

```

CDS
/genes="K07E3.4b"
complement(join(234. .394,582. .1530,1583. .1836,1896. .2352,
2402. .2500,3120. .3122))
/genes="K07E3.4b"
/notes="Contains similarity to Pfam domain: PF01268
(FTHFS), Score=1321.4, E-value=0, N=1; coded for by C.
elegans cDNA yk4f7.5; coded for by C. elegans cDNA
yk84g8.5; coded for by C. elegans cDNA yk87a4.5; coded for
by C. elegans cDNA yk25g7.5; coded for by C. elegans cDNA
yk45b11.5; coded for by C. elegans cDNA yk127d12.5; coded
for by C. elegans cDNA yk50e12.5; coded for by C. elegans
cDNA yk297h5.5; coded for by C. elegans cDNA yk25g7.3;
coded for by C. elegans cDNA yk15g9.5; coded for by C.
elegans cDNA yk45b11.3; coded for by C. elegans cDNA
yk50e12.3; coded for by C. elegans cDNA yk87a4.3; coded
for by C. elegans cDNA yk127d12.3; coded for by C. elegans
cDNA yk297h5.3; coded for by C. elegans cDNA yk12g12.5"
/codon_start=1
/protein_id="AAC69104.1"
/db_xref="GI:3805682"
/translation="MSTWNVYIKPTLLTPVPSDIAISRQPKDIEKVAIGEIKQE
EDLYGKAKVSLDILDRSEVKNKYVYVAGITPTPLGEGKSTTTIGVQALGAHL
HKVFCACVROSGPTFGIKGAAGGYSQVPIWEEFNHLTGDIHAITANLLAAA
IDARMHESQATDLNRLAPKNKGVRLSEIOLRRLDLGTIPRVDDAENLSEER
VSPARLNIOPATITWRVMDTNDRELFKEIGMGPNEKHTRTQFDITVASELMAL
ALTSIADMQERTARIYIGSDKAGNPVADDIGVTGALTVMRDTPVRLMQLGTP
VFVGFPAFNIAGHGSILADKVALKAGPDGVEITPAGFADIGMEKFNKCRYSG
LQPSAVLVATRALKMHGGSVAGAPLKHEYLNDINPLVEGGCDNSLRKQIENAN
KFGIPVVCVKFATDITDELELVSKAKYGAFAVVSHEWQSGGAGVALANSLVN
ATTGPKQFKFLNLDLSKIATIAKEIYAGDIELSDAKQKLERYTRQGFSLKP
ICWAKTHLSLSDPTKKGAPTFTPLRDVRSVAGGFYPLVGEMTTPGLNTRPCF
YDITDPVTEIIDLGL"
complement(join(234. .394,582. .1530,1583. .1836,1896. .2352,
2772. .2846))
/genes="K07E3.4"
/notes="Contains similarity to Pfam domain: PF01268
(FTHFS), Score=1277.4, E-value=0, N=1"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69101.1"
/db_xref="GI:1118055"
/translation="MSQFDMLSYISGPKICYFLINLSVAGEIGIKQEBLDLYGRKK
AKVSLDILDRSEVKNKYVYVAGITPTPLGEGKSTTTIGVQALGAHLHKVPACVR
QPSQGTFGIKGAAGGYSQVPIWEEFNHLTGDIHAITANLLAAAIDARFHD
TQATDLNRLAPKNKGVRLSEIOLRRLDLGTIPRVDDAENLSEERQVRSFARLND
PATITWRVMDTNDRELFKEIGMGPNEKHTRTQFDITVASELMALITSLADM
QERIARIVIGSDKAGNPVADDIGVTGALTVMRDTPVRLMQLGTPVYVHAGPTA
NIAGOSSILADKVALKAGPDGVEITPAGFADIGMEKFNKCRYSGLQPSAVLV
ATVRALKMHGGSVAGAPLKHEYLNDINPLVEGGCDNSLRKQIENANRFGIPVVC
VNFATDITDELELVSKAKYGAFAVVSHEWQSGGAGVALANSLVNATTHPKQF
KFLYNLDLSLEDKIATIAKEIYAGDIELSDAKQKLERYTRQGFSLKPTCMARKTHLS
LSDPTKKGAPTFTPLRDVRSVAGGFYPLVGEMTTPGLNTRPCFCDITIDPVT
EIIDGLF"
complement(3990. .5768)
/genes="K07E3.3"
complement(join(3990. .4216,4682. .5060,5183. .5356,
5646. .5768))
/genes="K07E3.3"
/notes="Contains similarity to Pfam domain: PF00763
(THF_DHG_CYH), Score=599.9, E-value=4.9e-177, N=1; coded
for by C. elegans cDNA cm08c3; coded for by C. elegans
cDNA yk60g7.3; coded for by C. elegans cDNA yk44h8.3;
coded for by C. elegans cDNA yk11f6.3; coded for by C.
elegans cDNA yk60g7.5; coded for by C. elegans cDNA
yk44h8.5; coded for by C. elegans cDNA yk11f6.5; coded
for by C. elegans cDNA yk15d3.3; coded for by C. elegans
cDNA yk184f3.3; coded for by C. elegans cDNA yk184f3.5"
/codon_start=1
/protein_id="AAC69100.1"
/db_xref="GI:1118056"
/translations="MVAEIVSGLEYSKKVLHDYQKIAKTRHEHPNFHVAIVQGN
RSDSNVYINSKKAKEIADGKLKLPDITITQGLAKREIMALNHDNEIDGIIQLPL
DCKHEIDADSVIDLIDPLDKVDGLTRINAGRLARGELQRTIFFCTPFGLYLVQOATG
```

```

DANFVSGKEVVLGRSKIVGSPAAALFLWHHTVTICHSKTPNKKERCLRADILIVAI
GRHEVKADIKGCAFVIDCGINVGDDPNRSKRIYGVDDTEAAKEVAGFELTPVPGVGVP
MTVAMLRNTEFEQAKRRRLGQKYK"
complement(6954. .14126)
/genes="K07E3.6"
complement(6954. .8550)
/genes="K07E3.8"
complement(join(6954. .7130,7347. .7598,7658. .7793,
10553. .10731,10778. .10990,11040. .11201,11244. .11525,
11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPLLEILIVERKSWIEEA
VNSDEEDQDFTVQPTVLRPPEQATVPENNEYSVIKQHPNSALORMSVIISTPSE
HSAHDMVFTKSGPEMIASLCIPDTIPEDYMEVVEYVQGRFLIAVASHVHFAK
ALKTPRIMSELEFLGLIYVNRKLDVTLSYNELSVANIRCVMTGDNLLTMSVA
REGIIPHTKAEFLITHSKTEKDPLKFLIKESVSSSENDITDSEIGAKVSMCGD
GANDCAALKAHAGISLSQAEASIAAPFTNPDIRCVPTVYKEGRCALVTSYAVSI
MAYLSNFEFLVNDGTNISDQFLYIDLITLVALFLPGLNTEASRLKSLGIPPRR
RLATSAFVSFVGOMFNITQTTGYLLVRQGSWVPNPELNDTMTTIGTVFTFTSC"
CMVLYAFVYSGKGPYRSYTNWLLGIIIFVIGAINVMVLETNMGFLMLMGVYVP
STMSVFLIATLSAGVFLSLLYEHFFVEKVVAHFESCTVMDLSSEFEFTMYDAVEL
VVVLGFFFLYLTSEQLPAPPKELAPMSMDTVEELRYDGVKNHILFGLNGTII
DVTRKGFYGPGRKAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWET
QKFKYLTVGRVLNKSSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(join(6954. .7130,7347. .7598,7658. .7785,
8511. .8550))
/genes="K07E3.8"
/notes="coded for by C. elegans cDNA yk295f5.5; coded for
by C. elegans cDNA yk295f5.5; coded for by C. elegans cDNA
yk47g2.5; coded for by C. elegans cDNA yk47g2.3"
/codon_start=1
/protein_id="AAD31555.1"
/db_xref="GI:4883517"
/translation="MYTVSVTLFHKSFMTDLSSWFFETMYDAVLVVLVGGFFYWL
RSEQPLPAPKELAPMSMDTVEELRYDGVKNHILFGLNGTIIYDVRKGFYGP
KAYGTLAGHDATRALGTMDQNAVSEWDDHTGISADEQETANEWETQFKYLTAVRL
VKNSEKADYGNKRSFVRGAESLDSIINGDDEGTTKDN"
complement(10549. .16404)
/genes="K07E3.7"
complement(join(10549. .10731,10778. .10990,11040. .11201,
11201. .11244. .11525,11759. .12052,12105. .12209,12257. .12395,12789. .13186,
13237. .13635,13968. .14126))
/genes="K07E3.6"
/notes="Contains similarity to Pfam domain: PF00122
(E1-E2_ATPase), Score=36.1, E-value=8.1e-10, N=4"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC69103.1"
/db_xref="GI:1118057"
/translation="MLNGESVPVTKASLKEADGCGPEIRLSSEHNRTLFSGTTVLQI
RNYKQPMIRILKTALESVNLKARVIRGFTSLKGLVRSIMYKPKQKEALKDV
MVFLVLGFTALIGFIYTVIEMVSRGSLKHIIRSLDITIVVPPALPAAMSGLIIN
ANSRLKKKICTSPTTNYVCGLINVACFDKGTGLTEDGLDFNCLKAIKRNDEKPEF
TSFEELDPVKLSAENANLIVAAASHLSRLTRIDGTLHGDPL
```

Search completed: April 19, 2000, 02:21:31  
Job time: 4677 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2000, 01:08:48 ; Search time 54.27 Seconds  
(without alignments)  
3632.787 Million cell updates/sec

Title: US-09-203-548-2

Perfect score: 788

Sequence: 1 GCCCGCAACCCGCGGCC.....AATATTAGAAAGTTTGAGC 788

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	784	99.5	788	1 V60245	Nucleotide sequenc
2	706.8	89.7	1875	1 V84368	Human stomach carc
3	679	86.2	1868	1 V63170	cDNA from clone bp
4	563	71.4	1893	1 X09852	Pig plasma membran
5	200.4	25.4	2776	1 X00678	Human secreted pro
6	42	5.3	2712	1 T94214	Polyangium brachys
7	42	5.3	2849	1 T13296	DNA contg. phae an
8	40.8	5.2	13144	1 Q13288	P.denitrificans ge
9	39.4	5.0	2241	1 V68070	Maize-optimised DN
10	39.4	5.0	2370	1 V68069	Maize-optimised DN
11	39.4	5.0	2403	1 T13955	Maize-optimised VI
12	39.4	5.0	2403	1 T74004	Maize-optimised B.
13	39.4	5.0	2403	1 V16177	Maize optimised DN
14	39.4	5.0	2403	1 V68065	Maize-optimised DN
15	39.4	5.0	2444	1 T13964	VIP3A(a) synthetic
16	39.4	5.0	2444	1 T74011	B. cereus VIP3A(a)
17	39.4	5.0	2444	1 V16190	DNA for maize opti
18	39.4	5.0	114955	1 X53491	Human adenosine A1
19	38.6	4.9	1260	1 V41730	Codon-optimised RA
20	38.4	4.9	2006	1 Q50147	Phospholipase D-P
21	38.4	4.9	7193	1 V50431	Streptomyces clavu
22	38	4.8	1044	1 V71739	Rice OSWAD56 cDNA.
23	37.6	4.8	1164	1 Q03159	Complete gene sequ
24	37.6	4.8	1459	1 T29043	Oerskovia beta-1,3
25	37.6	4.8	3435	1 T35869	Human DNA polymera
26	37.4	4.7	1618	1 T05503	Leishmania sp. ant
27	37.4	4.7	1618	1 T80384	DNA encoding Lheif
28	37.4	4.7	1618	1 V47559	Leishmania antigen
29	37.4	4.7	1618	1 X02202	L. braziliensis EI
30	37.4	4.7	4004	1 T51260	Human AD4 gene gen
31	37.2	4.7	936	1 T95102	Exon 3 encoding am
32	37.2	4.7	1867	1 T80385	DNA encoding Lheif
33	37.2	4.7	1951	1 V11459	C. acidivorans gam
34	37.2	4.7	2078	1 V41256	Human neuronal PAS

35 37 4.7 234 1 V15056 Xylanase activity  
36 37 4.7 348 1 T51709 Derivative #1 of a  
37 37 4.7 452 1 N91472 MPB-70 gene of Myc  
38 37 4.7 557 1 Q75446 Human glycosylatio  
39 37 4.7 831 1 Q99805 Thaumatin like gen  
40 37 4.7 906 1 Q05975 Sequence encoding  
41 37 4.7 3287 1 X26587 Nucleotide sequenc  
42 36.8 4.7 2094 1 T13378 Protease inhibitor  
43 36.6 4.6 1564 1 Q38955 Lipase gene #1. DN  
44 36.6 4.6 1620 1 Q22482 groEL-1 gene codin  
45 36.6 4.6 2668 1 Q22485 groEL-1 gene. Reco

## ALIGNMENTS

RESULT 1

V60245

ID V60245 standard; DNA; 788 BP.

AC V60245;

DT 14-DEC-1998 (first entry)

DE Nucleotide sequence encoding human cytokine/steroid receptor protein.

KW ss; human: cytokine/steroid receptor protein; CYSTAR;

KW developmental disorder; aberrant cellular differentiation; inflammation.

OS Homo sapiens.

FH key Location/Qualifiers

FT CDS 100..762

FT FT /\*tag= a

FT FT /product= "Cytokine/steroid receptor protein"

FT FT /transl\_except= (pos:280..282,aa:Xaa)

FT FT /transl\_except= (pos:709..711,aa:Xaa)

FT FT /transl\_except= (pos:712..714,aa:Xaa)

FT FT /transl\_except= (pos:745..747,aa:Xaa)

PN WO9841538-A2.

PD 24-SEP-1998.

PE 20-MAR-1998; U06045.

PR 20-MAR-1997; US-822264.

PA (INCY-) INCYTE PHARM INC.

PI Goli SK, Hillman JL, Murry LE;

DR WPI; 98-521162/44.

DR P-PSDB; W71362.

DR New cytokine-steroid receptor and related nucleic acid, vectors,

PT transformed cells - antibodies and antagonists, for diagnosis,

PT treatment and prevention of developmental disorders, abnormal

PT cellular differentiation and inflammation

PS Claim 1; Fig 1; 55pp; English.

CC The cytokine/steroid receptor protein designated CYSTAR. is used to treat

CC developmental disorders (e.g. spina bifida, hydrocephalus, renal tubular

CC acidosis, anaemia, congenital glaucoma or cataract and many others

CC disclosed). Antagonists are used to treat aberrant cellular

CC differentiation (e.g. hyper- or hypo-cortisolism, hyperaldosteronism, and

CC many disorders of the nervous, secretory, immune and haematopoietic

CC tissues) and inflammation (e.g. allergy, asthma, rheumatoid arthritis,

CC multiple sclerosis, osteoporosis, grave's disease, Crohn's disease and

CC many others disclosed).

CC Sequence 788 BP; 177 A; 219 C; 228 G; 160 T;

Query Match 99.5%; Score 784; DB 1; Length 788;

Best Local Similarity 100.0%; Pred. No. 3.2e-184;

Matches 788; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCGAACC CGCGCCACTCGCTCGAGAGGAGGAGAGAAAGTGCGAGTTCGG 60

Db 1 GCCGCGAACC CGCGCCACTCGCTCGAGAGGAGGAGAGAAAGTGCGAGTTCGG 60

QY 61 ATCCCTGCCCTAGCGGCGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGGTG 120

Db 61 ATCCCTGCCCTAGCGGCGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGGTG 120

QY 121 CGGACTGGCGCCGCGCCCAACCGATCTCGAGAGCGCGCGGCTGCTGCATGAGATTTTCACG 180

Db 121 CGGACTGGCGCCGCGCCCAACCGATCTCGAGAGCGCGCGGCTGCTGCATGAGATTTTCACG 180



Db 721 TCCACTCTCTCTTTAAACATAGTATTACATATATTAGAAAGTTT 766

## RESULT 3

V63170  
AC V63170 standard; cDNA; 1868 BP.

DE 12-JAN-1999 (first entry)  
CDNA from clone bp645\_10 which encodes a secreted protein.

KW Secreted protein; immune stimulating; suppressing;  
haematopoiesis regulating activity; tissue growth activity; activin;  
inhibin activity; chemotactic; chemokinetic activity; haemostatic;  
thrombolytic activity; anti-inflammatory activity; cadherin;  
tumour invasion suppressor activity; tumour inhibition activity; ds.  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT CDS 52..639  
ET /\*tag= a

PN WO98040486-A2.

PD 17-SEP-1998.

PE 13-MAR-1998; U04977.

PR 29-OCT-1997; US-960022.

PR 14-MAR-1997; US-815047.

PA (GENY ) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,

PI Racine LA, Spaulding V, Treacy M;

DR WPI; 98-520812/44.

DR P-PSDB; W80396.

PT New isolated human poly:nucleotide(s) and secreted proteins -

obtained from e.g. human foetal kidney, placenta, foetal brain,

adult testes, adult brain or adult uterus CDNA libraries

PS Claim 17; Pages 67-68; 110pp; English.

CC The present sequence encodes a secreted protein. The nucleic acid  
sequence is isolated from a human foetal kidney CDNA library using  
probe V63180. The polypeptide may have biological activities such as  
e.g. nutritional activity, immune stimulating or suppressing activity,  
haematopoiesis regulating activity, tissue growth activity,  
activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
activity or other activities.

SQ Sequence 1868 BP; 544 A; 390 C; 414 G; 518 T;

## Query Match

Best Local Similarity 86.2%; Score 679; DB 1; Length 1868;

Matches 718; Conservative 2; Mismatches 16; Indels 3; Gaps 3;

QY 49 GCGAGTTCGGATCCCTGCCTAGCGGGCCCAACCTTTACTCCAGAGATCATGGCTGCC 108

Db 1 GCGAGTTCGGATCCCTGCCTAGCGGGCCCAACCTTTACTCCAGAGATCATGGCTGCC 60

QY 109 GAGGATGTGTCGGAGTGGCGGACCCCAAGCGATCTGGAGAGCGGGCTGCTGCAT 168

Db 61 GAGGATGTGTCGGAGTGGCGGACCCCAAGCGATCTGGAGAGCGGGCTGCTGCAT 120

QY 169 GAGATTTTCACGTGCGCGTCAACCTGCTGCTGCTGGCCCTCTGCTGCTCTAC 228

Db 121 GAGATTTTCACGTGCGCGTCAACCTGCTGCTGCTGGCCCTCTGCTGCTCTAC 180

QY 229 AAGATGTCGGGGGACACCGCGGGCGGACGGCGGACGAGACGAGCGCGGCC 288

Db 181 AAGATGTCGGGGGACACCGCGGGCGGACGGCGGACGAGACGAGCGCGGCC 240

QY 289 CCTCTGCCCCGCTCAAGCGGCGGACTTACCCCCCGCGAGTGGCGGCTTCGACGGC 348

Db 241 CCTCTGCCCCGCTCAAGCGGCGGACTTACCCCCCGCGAGTGGCGGCTTCGACGGC 300

QY 349 GTCCAGGACCCCGCATACTCATGGGCATCAACGGCAAGGTGTTGCGATGTGACCAAGGC 408

Db 301 GTCCAGGACCCCGCATACTCATGGGCATCAACGGCAAGGTGTTGCGATGTGACCAAGGC 360

QY 409 CGCAAAATTCAGGGGCCCGAGGGGGCGGTATGGGGTCTTTGCTGGAAGAGATGATCCAGG 468

Db 361 CGCAAAATTCAGGGGCCCGAGGGCCGCTATGGGGTCTTTGCTGGAAGAGATGATCCAGG 420

QY 469 GGCCTTCCACATTTTCCCTGGGATAAGGAAGCACTGAAGATGAGTAGGACTTTCT 528

Db 421 GGCCTTCCACATTTTCCCTGGGATAAGGAAGCACTGAAGATGAGTAGGACTTTCT 480

QY 529 GACCTCACTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTTCACATTCAAG 588

Db 481 GACCTCACTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTTCACATTCAAG 540

QY 589 TATCATCACGTGGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTACTCAGATGAG 648

Db 541 TATCATCACGTGGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTACTCAGATGAG 600

QY 649 GAAGAACCAAAAGATGAGAGTTCGCCGAAAATG-TTAAAGCATTCAGTGGGAATATATC 707

Db 601 GAAGAACCAAAAGATGAGAGTTCGCCGAAAATG-TTAAAGCATTCAGTGGGAATATATC 660

QY 708 TA-TNNTGTATTTTGCACAAATCATTTTGAACAGTCCACTNTGCTTTTAAACATAGTG-T 765

Db 661 TATTTTGTATTTTGCACAAATCATTTTGAACAGTCCACTNTGCTTTTAAACATAGTG-T 720

QY 766 TACAATATTAGAAAGTTT 784

Db 721 TACAATATTAGAAAGTTT 739

## RESULT 4

V09852

ID V09852 standard; cDNA to mRNA; 1893 BP.

AC V09852;

DE 11-JUN-1998 (first entry)

KW Pig plasma membrane bound receptor cDNA.

KW opiod receptor; antibody; diagnosis; therapy; tumour cell;

OS Sus scrofa.

FH Key

FT CDS 42..626

ET /\*tag= a

ET /product= plasma\_membrane\_bound\_receptor

PN DB1962737-A1.

PD 08-JAN-1998.

PE 07-JUL-1996; 027237.

PR 07-JUL-1996; DE-027237.

PA (WEHL/) WEHLING M.

PI Wehling M;

DR WPI; 98-064316/07.

DR P-PSDB; W39900.

PT Nucleic acid encoding plasma membrane bound steroid or opiod

receptor - and related vectors, polypeptide(s) and antibodies, used

for diagnosis and therapy, particularly of tumour cells targetted by

ligands of receptor

PS Claim 4; Pages 11-13; 18pp; German.

CC The present sequence encodes a pig plasma membrane bound steroid or

opiod receptor, which can be used to generate antibodies for the

diagnosis and therapy of tumour cells in organs targetted by

steroid hormones and/or opiods, and nerve cells. Probes and

primers based on the receptor cDNA can be used to detect expression

of the receptor, and isolate related coding sequences. It can

also be used to produce recombinant polypeptides.

CC Sequence 1893 BP; 509 A; 447 C; 461 G; 476 T;

## Query Match

Best Local Similarity 71.4%; Score 563; DB 1; Length 1893;

Matches 652; Conservative 0; Mismatches 58; Indels 6; Gaps 5;

QY 72 GCGGGGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGAGTGTGGTGGGACTGGCGC 131

Db 14 GAGCCCTCCAACTTTTCTCCAGCGATCATGGCTGCCGAGAGTGTGGTGGGACTGGCGC 73





```

Db 673 ACCAAGGATCAACAAT 687
      || || || || || ||
RESULT 6
T94214
ID T94214 standard; DNA; 2712 BP.
AC T94214;
DT 21-MAY-1998 (first entry)
DE Polyangium brachysporum endoglucanase DNA.
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase; ss.
OS Polyangium brachysporum (clone 78Gal).
PN WO9744361-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08793.
PR 22-MAY-1996; US-651572.
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
PI Lam DE, Mathur EJ.
DR WPI: 98-018435/02.
DR P-PSDB; W35006.
PT Endoglucanase(s), preferably from archaeal bacterium, ABPII 1a -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
PS Claim 3; Fig IV; 164pp; English.
CC This DNA sequence from Polyangium brachysporum (clone 78Gal)
CC encodes an endoglucanase (see W35006) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic
CC bonds in cellulose, and which shows homology to the thermostable
CC endoglucanase (see W34985) of archaeobacterium hydrothermal vent
CC isolate ABPIIa. The DNA can be used in the recombinant production
CC of the endoglucanase and as a probe to identify similar sequences.
CC 24 Endoglucanase polynucleotides (see T94193-216) are claimed.
CC These can be incorporated into plasmid or virus-derived vectors for
CC use in a claimed method of producing enzymes in transformed host
CC cells. The claimed endoglucanases (see W34985-W35008) can be used
CC to degrade cellulose for the conversion of plant biomass into fuels
CC and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices.
SQ Sequence 2712 BP; 513 A; 919 C; 908 G; 372 T;

Query Match 5.3%; Score 42; DB 1; Length 2712;
Best Local Similarity 54.2%; Pred. No. 0.21;
Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 234 CGTGGCGGGGACCCGCGCGCCAGCGGCGGACAGCAGCAGCAGCGCGCCCTCT 293
      |||| || || || || || || || || || || || || || || || || ||
Db 381 CGTGACCGACTGGCTGGCGGTGGCGGACGCGCTTGTGGCGCGCTGATGATCCCGCAA 440
      |||| || || || || || || || || || || || || || || || || ||
QY 294 GCGCCGCCCTCAAGCGCGGAGCTTACCCCGCGGAGCTGGCGGCTTCACGCGCTCCA 353
      |||| || || || || || || || || || || || || || || || || ||
Db 441 GCGCGGGCGATCCCGCTCATCTGTGGGACCGCAGCGCGCTCCACGGCCAAACGTCAA 500
      |||| || || || || || || || || || || || || || || || || ||
QY 354 GGACCGCGCATACTCATGGCCATCAACGGAAGG 388
      || || || || || || || || || || || || || || || || || || ||
Db 501 GGGCGGACCATCTTCCCGCACAAACATCGGCTGG 535
      || || || || || || || || || || || || || || || || || || ||

RESULT 7
T13296
ID T13296 standard; DNA; 2849 BP.
AC T13296;
DT 17-OCT-1996 (first entry)
DE DNA contg. phaE and phac components of polyhydroxyacid synthase gene.
KW pha; polyhydroxyacid synthase; polyhydroxyacid production; Pseudomonas;
KW Alcaligenes; aerobic culture; biodegradable polymer; ss.
OS Thiocapsa pfennigii.
FH Key Location/Qualifiers
FT -35_signal 114..119
      /*tag= a

```

```

FT rbs 166..173
FT /*tag= b
FT /note= "Shine-Delgarno sequence"
FT cds 180..1283
FT /*tag= c
FT /note= "phaE"
FT rbs 1309..1317
FT /*tag= d
FT /note= "Shine-Delgarno sequence"
FT cds 1322..2395
FT /*tag= e
FT /note= "phaC"
PN DE4433134-A1.
PD 21-MAR-1996.
PF 16-SEP-1994; 433134.
PR 16-SEP-1994; DE-433134.
PA (BUCK-) BUCK WERKE GMBH & CO.
PA (MONS ) MONSANTO CO.
PI Liebergesell M, Pries A, Steinbuechel A, Valentin H;
DR WPI: 96-160943/17.
DR P-PSDB; R94501-02.
PT Polyhydroxy:acid prodn. using recombinant bacteria - esp.
PT Pseudomonas or Alcaligenes spp. contg. Thiocapsa polyhydroxy:acid
PT synthase gene
PS Claim 33; Page 23; 25pp; German.
CC The present sequence is a DNA sequence isolated from Thiocapsa pfennigii
CC contg. the phaE and phac components of the pha (polyhydroxyacid synthase)
CC gene. Recombinant bacteria contg. and expressing at least a part of the
CC pha gene are useful for the prodn. of polyhydroxy acids which are useful
CC as biodegradable polymers. The recombinant bacteria, e.g. Pseudomonas
CC putida Gp104(PHP1014::E156) and Alcaligenes eutrophus
CC PHB-4(PHP1013::E156), are cultured under aerobic conditions in a mineral
CC salts medium contg. a substrate carbon source selected from levulinic
CC acid, 5-hydroxyhexanoic acid, 4-hydroxyheptanoic acid, 4-hydroxyoctanoic
CC acid and their salts, esters, lactones and halogenated derivs.
SQ Sequence 2849 BP; 509 A; 1065 C; 878 G; 397 T;

Query Match 5.3%; Score 42; DB 1; Length 2849;
Best Local Similarity 49.3%; Pred. No. 0.21;
Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 145 CTGAGAGCGGGGCTGCTGATGAGATTTCAGTCGCGGCTCAACCTGCTGCTGCT 204
      |||| || || || || || || || || || || || || || || || || ||
Db 1373 CTGGAGTACAGCCGCAAGCTCGGCGAGGTATGAGAAGCTGCTCAAGCGCGACGATC 1432
      |||| || || || || || || || || || || || || || || || || ||
QY 205 GGCTCTGCTATCTCTGCTCTACAAAGATGCTGGCGGGGACGACGCGCGCGCGGC 264
      |||| || || || || || || || || || || || || || || || || ||
Db 1433 GACACAGGCTCACCCCAAGAGCTGCTCCACCGGAGGACAAGCTGCTCTTACCGC 1492
      |||| || || || || || || || || || || || || || || || || ||
QY 265 GACAGGACGACGACGCGCGCCCTCTGCGCGGCTCAAGCGCGGAGCTTCACGCC 324
      |||| || || || || || || || || || || || || || || || || ||
Db 1493 TACCGGCGCGCGGCGAGGTGGGACCCAGCAGATCCGCTGCTGATGCTGTACGCCCTC 1552
      |||| || || || || || || || || || || || || || || || || ||
QY 325 GCCGAGCTGGCGGCTTCGACGGGCTCCAGGACCGCGCGC 363
      |||| || || || || || || || || || || || || || || || || ||
Db 1553 GTCAATCGGCTTACATGACCGACATCCAGGAGGATCGC 1591
      |||| || || || || || || || || || || || || || || || || ||

RESULT 8
Q13288
ID Q13288 standard; DNA; 13144 BP.
AC Q13288;
DT 25-OCT-1991 (first entry)
DE P.denitrificans genes cob q, p, w, n and o.
KW cob gene; corrinoid; descobaltocorrinoid; cor gene; ds.
OS Pseudomonas denitrificans.
FH Key Location/Qualifiers
FT cds 429..1886
      /*tag= a
      /product= COBQ
      /note= "has cobyric acid synthase activity"
FT FT 3364..3888

```

```

FT FT      /*tag= b
FT FT      /product= COBP
FT FT      /note= "involved in conversion of cobinamide to
FT FT      GDP-cobinamide; has cobinamide kinase
FT FT      and cobinamide phosphate guanylyltransferase
FT FT      activity"
FT FT      3892..4956
FT FT      /tag= c
FT FT      /product= COBW
FT FT      5060..8887
FT FT      /tag= d
FT FT      /product= COBN
FT FT      9034..9678
FT FT      /tag= e
FT FT      /product= COBO
FT FT      /note= "has cob(II)alamin adenosyltransferase
FT FT      activity"
FT FT
FT FT      W09111518-A.
FT FT      08-AUG-1991.
FT FT      30-JAN-1991; F00054.
FT FT      31-JAN-1990; FR-001137.
FT FT      (RHON ) RHONE-POULENC BIOCH.
FT FT      Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;
FT FT      Thibaut D;
FT FT      WPI; 91-252650/34.
FT FT      P-PSDB: R13513-R13517.
FT FT      New polypeptide(s) involved in cobalamin and cobamide
FT FT      biosynthesis - and DNA encoding them, for amplification of
FT FT      cobalamin, esp. coenzyme B12 prodn.
FT FT      Claim 1; Fig 33; 299pp; French.
FT FT      This is a P.denitrificans genomic fragment isolated from a DNA bank
FT FT      constructed in vector pXL59. Plasmids able to complement P.putida and
FT FT      Agrobacterium tumefaciens cob mutants were identified and their
FT FT      inserts were sequenced. This insert was found to contain 5 putative
FT FT      open reading frames. The presence of cob genes n, o, p, q and w
FT FT      was subsequently verified by genetic analysis.
FT FT      See also Q13284-Q13287.
FT FT      Sequence 13144 BP; 2489 A; 4078 C; 4291 G; 2286 T;
FT FT
FT FT      Query Match          5.2%; Score 40.8; DB 1; Length 13144;
FT FT      Best Local Similarity 46.8%; Pred. No. 0.7;
FT FT      Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
FT FT
FT FT      QY 183 GCCGCTCAACCTGCTGCTGGCTCTGCATCTCTCTACAGATCGTGGCGG 242
FT FT      || ||||| || || ||||| ||||| || || ||||| || || ||||| ||
FT FT      DB 7006 GCTGCTACGCCGCCAGATCCTCGATCTCTGTCGCGACATCGCCCTCGACAGCGACG 7065
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      QY 243 GGACAGCGCGCGGCGGCGGACGAGCAGCAGCGCGCGCGCGCTCTGCCCGCGCT 302
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      DB 7066 CATGCACAGGGGCGACGCGGACGACAGGCGTGGAAGAAGCTCGAGCGCTATCTCTGCA 7125
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      QY 303 CAAGCGGCGGCACTTACCCCGCGGAGCTGGCGGCTTCGACGGCGTCCAGACCCGCG 362
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      DB 7126 CCTCAAGGAAATGCAGATCCGCGAGCGGCTGCACATCTTCGCGTTCGCGGGAAGGCG 7185
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      QY 363 CATACTCATGCGCATCACGGCAAGGTGTGATGTGACCAAGCGCGCAATTTACGG 422
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      DB 7186 GTTGTGACGACCTCACCGTAGCGCTGGCGGCTGCGGCCCGAGGTCTCGGCGAGGCGG 7245
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      QY 423 GCCCGAGGGCGGATGGGGTCTTTGCTG 451
FT FT      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
FT FT      DB 7246 CGACAGAGCTGCAGCGGGGATGCGAG 7274
FT FT
FT FT      RESULT 9
FT FT      V68070
FT FT      ID V68070 standard; DNA; 2241 BP.
FT FT      AC V68070;
FT FT      DT 02-FEB-1999 (first entry)
FT FT      DE Maize-optimised DNA encoding Bacillus VIP3A(c).
FT FT      KW VIP3A(c); toxin; apoptosis; biological control; insecticide;
FT FT      entomocide; plant pest; transgenic plant; maize; ss.

```

```

OS Bacillus thuringiensis.
OS Synthetic.
OS WO9844137-A2.
PD 08-OCT-1998.
PF 02-APR-1998; E01952.
PR 03-APR-1997; US-838219.
PR 03-APR-1997; US-832263.
PR 03-APR-1997; US-832265.
PA (NOVS ) NOVARTIS AG
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI Desai NM, Estruch JJ, Kozielec MG, Nye GJ, Warren GW,
PI Yu C;
PI WPI; 98-568281/48.
DR New VIP3A(c) protein and homologues active against plant pests -
PT used in entomocidal compositions for controlling insects and
PT arachnids, e.g. Lepidoptera and Coleoptera species
PT Disclosure: Page 85-86; 92pp; English.
PS This DNA sequence, encoding Bacillus thuringiensis VIP3A(c)
CC insecticidal protein (see W80322), has codons optimised for
CC expression in maize. The invention relates to VIP3A(c) and its
CC homologues. Also new are: (1) a protein comprising a toxic domain
CC of a VIP3 protein; (2) a transgenic plant containing DNA encoding a
CC VIP3 protein; (3) a microorganism containing a heterologous DNA
CC encoding VIP3A(a) (see W80320); (4) a recombinant DNA encoding
CC VIP3A(c) and its homologues; (5) an expression cassette comprising
CC a heterologous promoter operably linked to DNA encoding a VIP3A(c)
CC protein; (6) a receptor (see W80323) of the VIP3 class; (7) DNA
CC encoding the receptor of (6), and (h) antibodies to the receptor of
CC (6). The microorganisms are used in claimed entomocidal
CC compositions. Transgenic plants expressing VIP3A(c) are used to
CC control insects and arachnids, such as Coleoptera and Lepidoptera.
CC Sequence 2241 BP; 608 A; 698 C; 637 G; 298 T;
CC
SQ
Query Match          5.0%; Score 39.4; DB 1; Length 2241;
Best Local Similarity 46.9%; Pred. No. 0.86;
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 155 GCGGCTGCTGCATGAGATTTTCAGTCGCGCCTCAACCTGCTGCTGTGGCTCTGCA 214
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 344 GCGTGTACCTGCCGAAGATCACAGCATGCTGAGGAGCTGATGAGCAGAACTACGCC 403
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 215 TCTTCTCTCTACAGATCGTGGCGGACGACCGCGCGCGCGCGCGAGCTGC 274
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 404 TGAGCTGTCAGATCGAGTACCTTGAGCAAGCAGCTGCAGGAGATCAGCGAAGCTGG 463
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 275 CGACGANGCGCGCCCTCTGTCGCCCTCAAGCGCGGAGCTTACCCCGCGAGCTGC 334
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 464 TCATCAACGTGAACCTCTCTGATCAACAGCACCCCTGACCGAGATCACCCGCGCT 523
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 335 GCGGCTTCGAGCGGCTCCAGGACCGCGGCACTCATGCGCCATCAACGCAAGGTTCG 394
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 524 GCATCAAGTACCTGTAACGAGAAGTTCTGAAGAGCTGACCTTCGCCCCAGAGCAGCA 583
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 395 ATGTGACCAAGCGCGCA 412
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 584 AGGTGAAGAAGCAGCGCA 601
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
RESULT 10
V68069
ID V68069 standard; DNA; 2370 BP.
AC V68069;
DT 02-FEB-1999 (first entry)
DE Maize-optimised DNA encoding Bacillus VIP3A(b).
KW VIP3A(b); toxin; apoptosis; biological control; insecticide;
KW entomocide; plant pest; transgenic plant; maize; ss.
OS Bacillus thuringiensis.
OS Synthetic.
OS WO9844137-A2.
PD 08-OCT-1998.
PF 02-APR-1998; E01952.
PR 03-APR-1997; US-838219.

```



Best Local Similarity 46.9%; Pred. No. 0.88;  
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 155 GCGGCTGCTGCATGAGATTTTTCAGTGGCGCTCAACCTGCTGCTGTGCTTGCCTCTGCA 214  
DB 360 GCGTGTACCTGCCGAAGATCACCAGCATGCTGAGGAGCTGATGAAGCAGAACTACGCC 419  
QY 215 TCTTCTGCTCTACAAGATCGTGGCGGGGACACCGCGCCAGCGGCGGACGAGCA 274  
DB 420 TGAGCTGCATCGAGTACCTGAGCAGCAGCTGCAGAGATCAGCAGAGCTGGACA 479  
QY 275 CGACGANGCCGCCCTCTGCCCCGCTCAAGCGCGGAGCTTACCCCGCGGAGCTGC 334  
DB 480 TCATCAAGCTGAAGTCTCTGATCAACGACCCCTGACCGAGATCACCCGGCTACCAGC 539  
QY 335 GCGCTTCGAGCGGCTCCAGACCCCGCCATCTCATGCCCATCAACGGCAAGGTGTGC 394  
DB 540 GCATCAAGTACCTGAACGAGAGTTCGAAGAGCTGACCTTCGCCACCGAGACGAGCA 599  
QY 395 ATGTGACCAAGGCGCA 412  
DB 600 AGGTGAAGAGGCGCA 617

## RESULT 13

V16177  
ID V16177 standard; DNA; 2403 BP.  
AC V16177;  
DT 02-JUN-1998 (first entry)  
DE Maize optimised DNA sequence for VIP3A(a) protein.  
KW Vegetative insecticidal protein; VIP; expression; maize; protection;  
KW plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;  
KW recombinant; Bacillus thuringiensis; transgenic plant; resistance;  
KW insect attack; Sesamia; maize; cereal crop; ss.  
OS Synthetic.  
PN Bacillus sp.  
PS WO9746105-A1.  
PD 11-DEC-1997.  
PF 27-MAY-1997; E0237.  
PR 06-JUN-1996; GB-01177.  
PA (NOVS ) NOVARTIS AG.  
PI Hunter B, Suwantaradon K, Uitdewilligen WPM;  
DR WPI; 98-041787/04.  
PT Administration of Bacillus sp. toxin protein, especially Cry or  
PT vegetative insecticidal protein (VIP) protein to plants - useful for  
PT protection against attack by Asian Corn Borer (Ostrinia furnacalis)  
PS Claim 8; Pages 95-97; 175pp; English.  
CC The present sequence encodes a vegetative insecticidal protein 3A(a)  
CC (VIP3A(a)). The sequence is a synthetic sequence, in that the original  
CC codon usage of B. cereus has been optimised for expression in maize.  
CC VIP toxins and genes are used, especially inside recombinant B. cereus  
CC or B. thuringiensis strains, to produce plants protected against Asian  
CC Borer pests. Transgenic plants protected against Asian Corn Borer can  
CC be used to produce seed and progeny also resistant to insect attack.  
CC Plants expressing both a Cry-type and a VIP toxin gene can also protect  
CC against Sesamia pests. The method and compositions are especially used  
CC for protecting maize but may also be used to protect other cereal crops  
CC against Asian Corn Borer attack.  
SQ Sequence 2403 BP; 659 A; 743 C; 674 G; 327 T;

Query Match 5.0%; Score 39.4; DB 1; Length 2403;  
Best Local Similarity 46.9%; Pred. No. 0.88;  
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 155 GCGGCTGCTGCATGAGATTTTTCAGTGGCGCTCAACCTGCTGCTGTGCTTGCCTCTGCA 214  
DB 360 GCGTGTACCTGCCGAAGATCACCAGCATGCTGAGCGAGCTGATGAAGCAGAACTACGCC 419  
QY 215 TCTTCTGCTCTACAAGATCGTGGCGGGGACACCGCGCCAGCGGCGGACGAGCA 274  
DB 420 TGAGCTGCATCGAGTACCTGAGCAGCAGCTGCAGAGATCAGCAGAGCTGGACA 479

QY 275 CGACGANGCCGCCCTCTGCCCCGCTCAAGCGGCGGAGCTTACCCCGCGGAGCTGC 334  
DB 480 TCATCAAGCTGAACCTGCTGATCAACGACCCCTGACCGAGATCACCCGGCTACCAGC 539  
QY 335 GCGGCTTCGAGCGGCTCCAGGACCCCGCCATCTCATGCCCATCAACGGCAAGGTGTGC 394  
DB 540 GCATCAAGTACCTGAACGAGAGTTCGAAGAGCTGACCTTCGCCACCGAGACGAGCA 599  
QY 395 ATGTGACCAAGGCGCA 412  
DB 600 AGGTGAAGAGGCGCA 617

## RESULT 14

V58065  
ID V58065 standard; DNA; 2403 BP.  
AC V58065;  
DT 02-FEB-1999 (first entry)  
DE Maize-optimised DNA encoding Bacillus VIP3A(a).  
KW VIP3A(a); toxin; apoptosis; biological control; insecticide;  
KW entomocide; plant pest; transgenic plant; maize; ss.  
OS Bacillus thuringiensis.  
Synthetic.  
PN Key Location/Qualifiers  
FT CDS 11..2392 /\*tag= a  
FT FT  
PN WO9844137-A2.  
PD 08-OCT-1998.  
PF 02-APR-1998; E01952.  
PR 03-APR-1997; US-838219.  
PR 03-APR-1997; US-832263.  
PR 03-APR-1997; US-832265.  
PA (NOVS ) NOVARTIS AG.  
PI (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
PI Desai NM, Estruch JJ, Koziel MG, Nye GJ, Warren GW,  
PI Yu C;  
DR WPI; 98-568281/48.  
PT New VIP3A(c) protein and homologues active against plant pests -  
PT used in entomocidal compositions for controlling insects and  
PT arachnids, e.g. Lepidoptera and Coleoptera species  
PS Disclosure; Page 75-76; 92pp; English.  
CC This DNA sequence, encoding Bacillus thuringiensis VIP3A(a)  
CC insecticidal protein (see W80320), has codons optimised for  
CC expression in maize. The invention relates to a novel VIP3  
CC protein, designated VIP3A(c) (see W80322), and its homologues.  
CC Also new are: (1) a protein comprising a toxic domain of a VIP3  
CC protein; (2) a transgenic plant containing DNA encoding a VIP3  
CC protein; (3) a microorganism containing a heterologous DNA encoding  
CC a VIP3A(a) protein; (4) a recombinant DNA encoding a VIP3A(c)  
CC protein and its homologues; (5) an expression cassette comprising a  
CC heterologous promoter operably linked to DNA encoding a VIP3A(c)  
CC protein; (6) a receptor (see W80323) of the VIP3 class; (7) DNA  
CC encoding the receptor of (6), and (h) antibodies to the receptor of  
CC (6). The microorganisms are used in claimed entomocidal  
CC compositions. Transgenic plants expressing VIP3A(c) are used to  
CC control insects and arachnids, such as Coleoptera and Lepidoptera.  
SQ Sequence 2403 BP; 659 A; 743 C; 674 G; 327 T;

Query Match 5.0%; Score 39.4; DB 1; Length 2403;  
Best Local Similarity 46.9%; Pred. No. 0.88;  
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 155 GCGGCTGCTGCATGAGATTTTTCAGTGGCGCTCAACCTGCTGCTGTGCTTGCCTCTGCA 214  
DB 360 GCGTGTACCTGCCGAAGATCACCAGCATGCTGAGCGAGCTGATGAAGCAGAACTACGCC 419  
QY 215 TCTTCTGCTCTACAAGATCGTGGCGGGGACACCGCGCCAGCGGCGGACGAGCA 274  
DB 420 TGAGCTGCATCGAGTACCTGAGCAGCAGCTGCAGAGATCAGCAGAGCTGGACA 479  
QY 275 CGACGANGCCGCCCTCTGCCCCGCTCAAGCGGCGGAGCTTACCCCGCGGAGCTGC 334

```
Db 480 TCATCAAGCTGAACGCTCTGTATCAACAGCACCCCTGACCGAGATCACCCGGGCTTACCAGC 539
QY 335 GCGCTTCGACGGCTCCAGGACCGCGCATACTCATGCCATCAACGCGCAAGGTGTTCC 394
Db 540 GCATCAAGTACGTGACGAGAGGTTCCGAAGAGCTGACCTTCCGCCACCGAGACCAGCAGCA 599
QY 395 ATGTGACCAAAAGGCGCA 412
Db 600 AGGTGAAGAGGCGCA 617

RESULT 15
T13964
ID T13964 standard; DNA; 2444 BP.
AC T13964
DT 15-AUG-1996 (first entry)
DE VIP3A(a) synthetic:native gene fusion.
KW Pesticide; insecticide; biological control agent; Lepidoptera;
KW Coleoptera; transgenic plant; maize; insect resistance;
KW black cutworm; Agrotis ipsilon; VIP; ss.
OS Chimeric synthetic;
OC Chimeric Bacillus thuringiensis strain AB88 (NRRL B-21225).
FH Key
FH Location/Qualifiers
FH cds 17..2444
FT /*tag= a
FT misc_difference 2384
FT /*tag= b
FT /*note= "base n at position 2384 is not identified
FT in the specification"
FT misc_difference 2420
FT /*tag= c
FT /*note= "base n at position 2420 is not identified
FT in the specification"
FT
FT WO9610083-A1.
PN 04-APR-1996.
PD
PF 27-SEP-1995; E03826.
PR 28-SEP-1994; US-314594.
PR 03-JUN-1995; US-463483.
PA (CIBA ) CIBA GEIGY AG.
PI Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
PI Kozziel MG, Mullinsma, Nye GJ, Warren GW;
DR WPI: 96-200921/20.
DR P-PSDB: R91261.
PT Bacillus strain producing insecticidal protein during vegetative
PT growth - used in the control of Lepidoptera and Coleoptera pests
PS Disclosure: Page 217-220; 242pp; English.
CC A VIP3A(a) gene (T13964) codes for a synthetic:native fusion
CC (R91261) of AB88 VIP3A(a), an insect-specific protein of Bacillus
CC thuringiensis AB88 that shows activity against Agrotis ipsilon,
CC Spodoptera frugiperda, Heliothis virescens and Helicoverpa zea.
CC Sequence 2444 BP; 694 A; 698 C; 654 G; 396 T;
SQ
```

```
Query Match 5.0%; Score 39.4; DB 1; Length 2444;
Best Local Similarity 46.9%; Pred. No. 0.89;
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 155 GCGGGCTGCTCATGAGATTTTCAGCTGCGCGCTCAACCTGCTGCTTGGCTCTGCA 214
Db 360 GCGTGTACTCGCGAGATCACCAGCATGCTGAGGAGCGTGATGAGAGACTAGGCC 419
QY 215 TCTTCTGCTCTACAGATCGTGGCGGGACCAAGCGCGGCCGCGAGAGACGA 274
Db 420 TGAGCTGCAGATCAGTACCTGAGCAAGCAGCTGAGGAGATCAGCGACAAGCTGGACA 479
QY 275 CGAGGANGCGCCCTCTGCGCCGCTCAAGCGGGGACTTACCCCGCGAGCTGC 334
Db 480 TCATCAAGCTGAACGCTCCTGATCAACAGCACCTTACCGAGATCACCCGGCTTACCAGC 539
QY 335 GCGGCTTCGAGCGGCTCCAGGACCGCGCATACTATGCGCATCAACGCGCAAGGTGTTCC 394
Db 540 GCATCAAGTACGTGACGAGAGGTTCCGAAGAGCTGACCTTCCGCCACCGAGACCAGCAGCA 599
```

```
QY 395 ATGTGACCAAAAGGCGCA 412
Db 600 AGGTGAAGAGGCGCA 617
```

Search completed: April 19, 2000, 02:20:16  
Job time: 4288 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2000, 01:07:02 ; Search time 37.32 Seconds  
(without alignments)  
2527.916 Million cell updates/sec

Title: US-09-203-548-2  
Perfect score: 788  
Sequence: 1 GCCCGCAACCCGCGGCC.....AATATTAGAAAGTTTGAGC 788

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	679	86.2	1868	4	US-08-960-022-5		Sequence 5, Appli
2	42	5.3	2849	5	US-08-809-286B-1		Sequence 1, Appli
3	41.2	5.2	2712	4	US-08-410-784A-1		Sequence 1, Appli
4	39.4	5.0	2241	3	US-08-838-219B-20		Sequence 20, Appl
5	39.4	5.0	2370	3	US-08-838-219B-19		Sequence 19, Appl
6	39.4	5.0	2403	2	US-08-471-033-30		Sequence 30, Appl
7	39.4	5.0	2403	3	US-08-471-044-30		Sequence 30, Appl
8	39.4	5.0	2403	3	US-08-463-483A-30		Sequence 30, Appl
9	39.4	5.0	2403	3	US-08-471-046A-30		Sequence 30, Appl
10	39.4	5.0	2403	3	US-08-470-566B-30		Sequence 30, Appl
11	39.4	5.0	2403	3	US-08-838-219B-7		Sequence 7, Appli
12	39.4	5.0	2403	4	US-08-469-334-30		Sequence 30, Appl
13	38.4	4.9	2523	4	US-08-410-784A-3		Sequence 3, Appli
14	38	4.8	1043	4	US-08-867-087B-12		Sequence 12, Appl
15	38	4.8	4766	4	US-08-852-401-1		Sequence 1, Appli
16	37.6	4.8	1164	1	US-07-640-476-6		Sequence 6, Appli
17	37.6	4.8	1459	4	US-08-824-707-1		Sequence 1, Appli
18	37.6	4.8	3435	2	US-08-366-577-1		Sequence 1, Appli
19	37.6	4.8	3435	6	PCT-US96-00005-1		Sequence 1, Appli
20	37.4	4.7	1618	3	US-08-533-669A-9		Sequence 9, Appli
21	37.4	4.7	1618	3	US-08-607-509-1		Sequence 1, Appli
22	37.4	4.7	1618	3	US-08-454-036-1		Sequence 1, Appli
23	37.4	4.7	1618	3	US-08-634-642-1		Sequence 1, Appli
24	37.4	4.7	1618	6	PCT-US95-05064-1		Sequence 1, Appli
25	37.4	4.7	2312	1	US-07-736-178C-1		Sequence 1, Appli
26	37.2	4.7	936	1	US-08-018-977C-4		Sequence 4, Appli
27	37.2	4.7	1867	3	US-08-607-509-3		Sequence 3, Appli

28	37.2	4.7	1867	3	US-08-634-642-3	Sequence 3, Appli
29	37.2	4.7	2078	3	US-08-785-310A-1	Sequence 1, Appli
30	37	4.7	348	4	US-08-610-728B-2	Sequence 2, Appli
31	37	4.7	557	2	US-08-416-336-7	Sequence 7, Appli
32	37	4.7	557	6	PCT-US94-05354-37	Sequence 37, Appl
33	36.6	4.6	1620	3	US-08-461-775-10	Sequence 10, Appl
34	36.6	4.6	2668	3	US-08-461-775-11	Sequence 11, Appl
35	36.2	4.6	1998	7	5212296-8	Patent No. 5212296
36	36.2	4.6	3489	4	US-08-728-323A-1	Sequence 1, Appli
37	36.2	4.6	32207	3	US-08-770-379-20	Sequence 20, Appl
38	36	4.6	1401	4	US-08-812-412-1	Sequence 1, Appli
39	36	4.6	2277	2	US-08-676-967-2	Sequence 2, Appli
40	36	4.6	2277	2	US-08-676-974-2	Sequence 2, Appli
41	36	4.6	2277	4	US-09-098-487-2	Sequence 2, Appli
42	35.6	4.5	3090	6	PCT-US93-06251-7	Sequence 7, Appli
43	35.6	4.5	4112	2	US-08-340-203A-2	Sequence 2, Appli
44	35.6	4.5	4112	3	US-08-452-567-2	Sequence 2, Appli
45	35.6	4.5	4112	4	US-08-452-427-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-960-022-5  
; Sequence 5, Application US/08960022  
; Patent No. 5978837  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/960,022  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1868 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-960-022-5

Query Match 86.2%; Score 679; DB 4; Length 1868;  
Best Local Similarity 97.2%; Pred. No. 4.4e-167;

	Matches	718;	Conservative	2;	Mismatches	16;	Indels	3;	Gaps	3;
Qy	49	GGCGAGTTC	CGGATCCCT	CGCTAGCGCGGCC	CCAACTTT	TACTCCAGAGAT	CATGGCTGCC	108		
Db	1	GGCGAGTTC	CGGATCCCT	CGCTAGCGCGGCC	CCAACTTT	TACTCCAGAGAT	CATGGCTGCC	60		
Qy	109	GAGGATGTG	TGGGACTT	GGCGCGAC	CCCAAGCGAT	CTGGAGAGCG	GCGGCTGCTGCAT	168		
Db	61	GAGGATGTG	TGGGACTT	GGCGCGAC	CCCAAGCGAT	CTGGAGAGCG	GCGGCTGCTGCAT	120		
Qy	169	GAGATTTTC	AGTCGCC	CGCTCAACT	GTGCTGTG	CTGTGGCCT	CTGTGATCTT	228		
Db	121	GAGATTTTC	AGTCGCC	CGCTCAACT	GTGCTGTG	CTGTGGCCT	CTGTGATCTT	180		
Qy	229	AAGATCTG	CGCGGGG	ACCGCGCG	CGCCAGCGCG	CACAGGAC	GACGAGCGCGGCC	288		
Db	181	AAGATCTG	CGCGGGG	ACCGCGCG	CGCCAGCGCG	CACAGGAC	GACGAGCGCGGCC	240		
Qy	289	CCTCTGCC	CGCCCTCA	AGCGCGG	CGACTTCA	CCCCCGCG	AGCTCGCGGCTTCGACGGC	348		
Db	241	CCTCTGCC	CGCCCTCA	AGCGCGG	CGACTTCA	CCCCCGCG	AGCTCGCGGCTTCGACGGC	300		
Qy	349	GTCCAGGAC	CCCGCAT	ACTCAT	GCCCATCA	ACGGCAAG	GTTCGATGTGACCAAGGC	408		
Db	301	GTCCAGGAC	CCCGCAT	ACTCAT	GCCCATCA	ACGGCAAG	GTTCGATGTGACCAAGGC	360		
Qy	409	CGCAAA	TTCTAC	GGGCGCG	AGGGCCG	TATGGG	GTCTTGTGGAAGAGATGCATCCAGG	468		
Db	361	CGCAAA	TTCTAC	GGGCGCG	AGGGCCG	TATGGG	GTCTTGTGGAAGAGATGCATCCAGG	420		
Qy	469	GGCCTTG	CCACAT	TTTGGCT	CGATGA	AGAAAG	CACTGAGATGACGTACCTTTCT	528		
Db	421	GGCCTTG	CCACAT	TTTGGCT	CGATGA	AGAAAG	CACTGAGATGACGTACCTTTCT	480		
Qy	529	GACCTCA	CTGCTGCC	ACGAGAC	TCTGAG	TGACTGGG	AGTCTCAGTTTCAAG	588		
Db	481	GACCTCA	CTGCTGCC	ACGAGAC	TCTGAG	TGACTGGG	AGTCTCAGTTTCAAG	540		
Qy	589	TATCAT	CACGTGGG	CAAACT	GCTGAAG	GAGGGG	GAGGCCAC	CTGTCTCTCAGATGAG	648	
Db	541	TATCAT	CACGTGGG	CAAACT	GCTGAAG	GAGGGG	GAGGCCAC	CTGTCTCTCAGATGAG	600	
Qy	649	GAAGAAC	CAAAAG	ATGAGAG	TTCCCG	AAAAATG	-TTAAAGCATTCAGTGGAAAGTATATC	707		
Db	601	GAAGAAC	CAAAAG	ATGAGAG	TTCCCG	AAAAATG	-TTAAAGCATTCAGTGGAAAGTATATC	660		
Qy	708	TA-TNNT	GATATT	TGCAAA	ATCATTT	GTACAG	TCCACCTNTGCTCTTTAAAAACATAGTG-T	765		
Db	661	TATTTT	TGATTT	TGCAAA	ATCATTT	GTACAG	TCCACCTNTGCTCTTTAAAAACATAGTGAT	720		
Qy	766	TACAAT	ATTTAGA	AGTTT				784		
			</							

```

RESULT      2
US-08-809-286B-1
; Sequence 1, Application US/08809286B
; Patent No. 6011144
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for manufacturing polyhydroxylic
; TITLE OF INVENTION: fatty acids, and recombinant bacterial strains for
; TITLE OF INVENTION: carrying out the process
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,286B
; FILING DATE: 3-JUL-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patricia A. Kammerer
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:152 (28-21(15115)A)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2849 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thiocapsa pfennigii
; IMMEDIATE SOURCE:
; CLONE: Pseudomonas putida SK 6691/Alcaligenes
; CLONE: eutrophus SK 6891
; US-08-809-286B-1
;
Query Match 5.3%; Score 42; DB 5; Length 2849;
Best Local Similarity 49.3%; Pred. No. 0.05;
Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 145 CTGAGAGCGGGCGCTGCTCATGAGATTTACGTCGCGCTCAACCTGCTGCTGCTT 204
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1373 CTGGAGTACAGCCGACAGCTGCGGGGGTATGAGAACCTGCTCAAGGCGCAGATC 1432
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 205 GGCCTCTGCATCTTCCTGCTCTACAAGATCGTGGCGGGGACGCGGCGCAGCGGC 264
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1433 GACACAGCGTCACCCCCAGGAGCGTGTCTACCGCGGAGACAAGCTGTGTCCTTACGCG 1492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 265 GACAGGACGAGCAGCGCGCCCTCTGCGCCCGCTCAAGCGGCGGACTTCACCCCC 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1493 TACCGGCGCCGCGCAGGTGGCGACCCACGACGATCCGGTGTGATCGTCTACGCCCTC 1552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 325 GCCGAGTGGGGCGCTTCGACGGCGTCCAGAACCCGCGC 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1553 GTCAATCGGCCCTACATGACCGACATCCAGGAGGATCGC 1591
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT      3
US-08-410-784A-1/c
; Sequence 1, Application US/08410784A
; Patent No. 5912413
; GENERAL INFORMATION:
; APPLICANT: MYERS, ALAN M.
; APPLICANT: JAMES, MARTHA G.
; TITLE OF INVENTION: ISOLATION OF SUL, A STARCH DEBRANCHING
; TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
; TITLE OF INVENTION: SUGARY 1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```









```

:
: NUMBER OF SEQUENCES: 50
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NV
: COUNTRY: USA
: ZIP: 10532
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/463,483A
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Spruill, W. Murray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8615
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2403 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: HYPOTHETICAL: NO
: FEATURE:
:
: NAME/KEY: misc_feature
: LOCATION: 11..2389
: OTHER INFORMATION: /note= "maize optimized DNA
:
: US-08-463-483A-30
: sequence encoding Vip3a(a)"
:

```

Query Match	5.08;	Score 39.4;	DB 3;	Length 2403;
Best Local Similarity	46.9%;	Pred. No. 0.22;		
Matches 121;	Conservative 0;	Mismatches 137;	Indels 0;	Gaps 0;
QY 155	GCGGGCTGCTCCATGAGATTTTCAGTGC	CGCGCTCAACCTGCTGCTGTGGCCCTCTGCA	214	
Db 360	GCGTGTACCTCCGAAGATCACCA	GATGCTGAGCGAGCTGATGAAGCAAACTACGCCC	419	
QY 215	TCATTCTGCTCTACAAGATCGTGGCGGGGACCAGCCGCGCCAGCGCGCAGCAGCA	274		
Db 420	TGAGCCTGCAGATCGATGATCTGAGCAAGCAGCTGCAAGGAGATCAACGACAAAGCTTGGACA	479		
QY 275	CGACGANGCCGCCCTCTGCGCCGCTCAAGCGCGCGACTTCAACCCCGCCGAGCTGC	334		
Db 480	TCATCAAGTGAACGTCTGTATCAACAGCACCTTGACCGAGATCAACCCGGCGCTACCAAGC	539		
QY 335	GGCGCTTCGAGCGGTTCAGGACCGCGCATCTCATGGCCATCAACGGCAAGGTGTTCG	394		
Db 540	GCATCAAGTACGTGAACGAGAAGTTCGAAGAGCTGACCTTCGCCACCGACCCAGCAGCA	599		
QY 395	ATGTGACCAAGGCCGA	412		
Db 600	AGGTGAAGAGGACGGCA	617		

Query Match 5.0%; Score 39.4; DB 3; Length 2403;  
Best Local Similarity 46.9%; pred. No. 0.22;

Matches	121;	Conservative	0;	Mismatches	137;	Indels	0;	Gaps	0;
Oy	155	GC GGCGTCTG CATGAGATTTTCACGTCCGGCTCAACCTGCTGCTTGCCCTCTGCA	214						
Db	360	CGGTGTACTCTGCCGAAGATCACCAAGCATGCTTGAGCGAGCTGATGAAGCAGAATACGCCCC	419						
Oy	215	TCTTCTGCTCTCAAAAGATCGTGC GGGGACCAGCGCGCGCCAGCGCGGCACAGGACGA	274						
Db	420	TGACCTTGCAGATCGATGATTACCTTGACAAGCAGCTGCAGAGATCAGCCACAAGCTGGACA	479						
Oy	275	CGACGANCCGCCCTCTTGCCCGCCCTCAAAGCGCGCGACTTTCACCCCCC CGCAGCTGC	334						
Db	480	TCATCAACGTAACGTCTGTATCAACAGCACCTTGACCGAGATCACCCCGGCTTACCAGC	539						
Oy	335	GGCGCTTCGAGGGCGTCCAGGACCCGCGCATATCATGCCATCAACCGCAGAGGTGTCG	394						
Db	540	GCATCAAGTAGCTGAACGAGAAGTTTCGAAGAGCTGACCTTGCCACCCGAGACAGCAGCA	599						
Oy	395	ATGTGACCAAGGCCGA	412						
Db	600	AGGTGAAGAGGAGCGCA	617						

```

RESULT 10
US-08-470-566B-30
: Sequence 30, Application US/08470566B
: Patent No. 5873212
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5872212artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,566B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,483
: FILING DATE: 05-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 36,241
: REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587

```

```

; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHEITICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3A(a)"
; US-08-470-566B-30

Query Match          5.0%; Score 39.4; DB 3; Length 2403;
Best Local Similarity 46.9%; Pred. No. 0.22;
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 155 GCGGGCTGCTCATGATGATTTTACGTCGGCGCTCAACCTGCTGCTGTTGGCCTCTGCA 214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 GCGTGTACCTGCCGAAGATCACCAGCATGCTGTAGCGACGTGTGAACAGCACTACGCC 419
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 215 TCTTCCTGCTCTCAAGATCGTGGCGGGGACCAAGCGGGCGGACAGGACAGGACGA 274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 TGAGCCTGCAGATCGAGTACTCTGAGCAAGCAGCTGCAAGGAGATCAGGCAAGGTGACA 479
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 275 CGAGGANGCGGCCCTCTGCCCGGCTCAAGCGGGCGGACTTCACCCCCCGCGAGTCG 334
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 TCATCAACGTGAAGTCCTCTGATCAACAGCACCTTGACCGGAGATCAACCCGGCTTACCAGC 539
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 335 GGGCCTTCGACGGGGCTCCAGGACCGCGGCATCTATTGGGCATCAACGGCAAGGTGTG 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 540 GCATCAAGTACGTGAACGAGAAGTTCGAAGAGCTGACCTTCGCCACCGACGACGACGA 599
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 395 ATGTGACCAAGGCCGA 412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 AGGTGAAGAAGGACGGA 617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 11
US-08-838-219B-7
; Sequence 7, Application US/08838219B
; Patent No. 5877012
;
; GENERAL INFORMATION:
;   APPLICANT: Warren, Gregory W
;   APPLICANT: Koziel, Michael G
;   APPLICANT: Mullins, Martha A
;   APPLICANT: Nye, Gordon J
;   APPLICANT: Carr, Brian
;   APPLICANT: Desai, Nalini M
;   APPLICANT: Kostichka, N. Kristy
;   APPLICANT: Duck, Nicholas B
;   APPLICANT: Estruch, Juan J
;
; TITLE OF INVENTION: A No. 5877012el Class of Proteins for the
;   TITLE OF INVENTION: Control of Plant Pests
;
; NUMBER OF SEQUENCE: 20
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: CIBA-GEIGY Corporation
;   STREET: 7 Skyline Drive
;   CITY: Hawthorne
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 10532
;
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30B
;   CURRENT APPLICATION DATA:

```





US-08-852-401-1  
; Sequence 1, Application US/08852401  
; Patent No. 5976836  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Larsen, Peter E.  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Methods and Compositions for Enhancing  
; TITLE OF INVENTION: Erythromycin Production  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
; STREET: 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,401  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa L.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: PER2159POO30US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4776 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 4.8%; Score 38; DB 4; Length 4776;  
Best Local Similarity 46.1%; Pred. No. 0.67;  
Matches 125; Conservative 0; Mismatches 146; Indels 0; Gaps 0;  
QY 117 GTTGGCGACTGGCGCGGACCCAGCGATCTGGAGAGCGGCGGCTGCTGATGAGATTTT 176  
Db 3802 GGTGGCCACAGAGGGGACACACCTACGGGCGATGGGTGAGATCGACACCGGTGCCG 3861  
QY 177 CAGTCGCGCGCTCAACTGCTGCTGTGGCTCTGCTGCTGCTCTACAGATCGT 236  
Db 3862 CTGGACCCCGCTACCGGACCTGACACCGACACCGTCTCGGTGCGCCAGCAGGTCCG 3921  
QY 237 GCGCGGGGACCGCGCGGCCAGCGCGGACAGGACGACGACGACGACGACGACGACGACG 296  
Db 3922 GCTCGCGCGCAAGATCGGACCGCGCGGTCGCGCGGCGGTGACCGGCGCGCCACCTGCA 3981  
QY 297 CGGCCTCAAGCGCGCGGACTTACCCCGCGCGGCTTCCAGCGCGTTCAGCGCGTCCAGGA 356  
Db 3982 CTACGAACAGCGGCTCAAGCGGACCGCGCAGAGCAAGCTCAACGCGCTCGCGGTCCC 4041  
QY 357 CCGCGCATACTGTCGATCAACGCGCAAG 387  
Db 4042 GTACTAGCGGCACCGCGCTTCCAGCAAG 4072

Search completed: April 19, 2000, 02:19:08  
Job time: 4326 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2000, 01:01:33 ; Search time 287.25 Seconds  
(without alignments)  
10357.609 Million cell updates/sec

Title: US-09-203-548-2  
Perfect score: 788  
Sequence: 1 GCGCGCGACCCCGCGGCC.....AATATTAGAAAGTTTGAGC 788

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues 9077268  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:  
1: em\_est11:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	541	68.7	720	47	AI479776
2	454.2	57.6	531	33	AA428520

AI479776 tm69c04.x  
AA428520 zw47c07.r

```

c 3 440 55.8 567 45 A1394728
c 4 437 55.5 478 64 A1035776
5 430.4 54.6 520 26 W56474
6 424.8 53.9 534 39 AA848529
7 416 52.8 535 29 AA184910
8 411.6 52.2 501 28 AA081900
c 9 407.6 51.7 573 62 A1885620
10 398 50.5 501 27 AA023630
11 383 48.6 508 31 AA288129
12 381.4 48.4 528 30 AA274818
13 380.8 48.3 512 28 AA117455
14 373 47.3 429 28 AA101294
c 15 371.4 47.1 458 25 N52291
c 16 367.4 46.6 525 46 A1453643
c 17 349.2 44.3 464 45 A1334835
18 345.4 43.8 476 28 AA111285
19 345.2 43.8 390 29 AA166645
20 340.8 43.2 438 33 AA389446
c 21 339.2 43.0 464 48 A1575080
22 335 41.2 429 35 A572534
23 318 40.4 369 22 RS2821
24 308.4 39.1 411 24 H95223
c 25 306.2 38.9 429 37 AA699865
26 298.6 37.9 438 23 H23045
27 297.6 37.8 406 28 A115422
c 28 288.4 36.6 438 27 AA021062
29 287.4 36.5 411 25 N69942
30 285.2 36.2 381 28 AA088127
31 284 36.0 401 26 W97563
c 32 283.6 36.0 349 30 AA232394
33 275.8 35.0 446 30 Z84149
34 273.6 34.7 359 25 W12485
35 272.6 34.6 367 27 AA016712
36 272 34.5 352 28 AA106719
37 270.4 34.3 498 31 AA316201
c 38 270 34.3 344 25 N58287
39 266 33.8 330 28 AA080939
c 40 255.4 32.4 334 23 H48290
41 251.2 31.9 314 26 W25549
c 42 241.2 30.6 461 69 AW132146
c 43 235.4 29.9 284 23 H60595
c 44 230.8 29.3 300 61 A1843194
c 45 230.8 29.3 300 69 AW125087

```

## ALIGNMENTS

```

RESULT 1
LOCUS A1479776/c
DEFINITION A1479776 720 bp mRNA EST 14-APR-1999
tm69c04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163366 3'
similar to TR:000264 000264 PUTATIVE PROGESTERONE BINDING PROTEIN.
; mRNA sequence.
ACCESSION A1479776
VERSION A1479776.1 GI:4372944
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 720)
AUTHORS NCI/NIHNS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/PTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188831.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

```

```

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 610 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 469.
FEATURES
Location/Qualifiers
Source
1..720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGGCGGCATAGTGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 166 a 191 c 179 g 180 t 4 others
ORIGIN

Query Match 68.7%; Score 541; DB 47; Length 720;
Best Local Similarity 94.8%; Pred. No. 3.5e-115;
Matches 619; Conservative 0; Mismatches 28; Indels 6; Gaps 5;

QY 136 CCAAGCGATCTGGAGCGCGGGCTGCTGCATGAGATTTTCAGTCGCGCTCAACTG 195
Db 720 CCAAGCGATCTGGAGCGCGGGCTGCTGCATGAGA-TTTCAGTCGCGCTCAACTG 662
QY 196 CTGCTGCTTGGCTCTGCATCTTCTCTCTCTACAAAGATCGTGCAGCGGACCGCGCG 255
Db 561 TNGCTGCTTGGCTCTGCATCTTCTCTCTCTACAAAGATCGTGCAGCGGACCGCGCG 602
QY 256 GCC-AGCGCGACAGGACGACGACGANGCGCGCCCTCTGCCCGCTCAAGCGCGCGA 314
Db 601 GCCAAGCGCGACAGGACGACGACGACGCGCGCCCTCTGCCCGCTCAAGCGCGCGA 542
QY 315 CTTACCCCGCGCGAGCTGGCGCTTCGAGCGGCTCCAGGACCGCGCATCTATGCG 374
Db 541 CTT-ACCCCGCGCGAGCTGGCGCTTCGAGCGGCTTCAGGACCGCGCATCTATGCG 483
QY 375 CATCAACGGCAAGGTGTCATGTGACCAAGCGCGCAAAATTCACGGCGCGAGGGGCC 434
Db 482 CATCAACGGCAAGGTGTCATGTGACCAAGCGCGCAAAATTCACGGCGCGAGGGGCC 423
QY 435 GTATGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTGCACATTTTGCCTGGATAA 494
Db 422 GTATGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTGCACATTTTGCCTGGATAA 363
QY 495 GGAAGCACTAGAGATGAGTACGATGACCTTCTGACCTCCTGCTCCAGCAGGAGAC 554
Db 362 GGAAGCACTAGAGATGAGTACGATGACCTTCTGACCTCCTGCTCCAGCAGGAGAC 303
QY 555 TCTGAGTACTGGGAGTCTCAGTTCACTTTCAAGTATCATCACTGCGGCAAACTGCTGAA 614
Db 302 TCTGAGTACTGGGAGTCTCAGTTCACTTTCAAGTATCATCACTGCGGCAAACTGCTGAA 243
QY 615 GGAGGGGAGGAGCGCCACTGTGTACTCAGATGAGAGAGAACCAAAAGATGAGATTCGCC 674
Db 242 GGAGGGGAGGAGCGCCACTGTGTACTCAGATGAGAGAGAACCAAAAGATGAGATTCGCC 183
QY 675 GAAAAATG-TTAAAGCATTCAGTGGAGATATATCTA-TNNTGTATTTTGGAAAAATCATTT 732

```

QY	129	CGCGACCCCAAGCGATCTGGAGACGCGGGGCGTGTGTCATGAGATTTTCAAGTCGCCGCT	188
Db	121	CG-CGACCCAAGCGATCTGGAGAGCGCGGCTGTGTCATGAGATTTTCAAGTCGCCGCT	179
QY	189	CAACTCTGCTGCTTGCCCTCTCATCTTCTCTACAGATCGTGC GGGGACCA	248
Db	180	C A C C T G C T G T G T T G G C C T C T C A T C T T C T A C A A G A T C G T G C G G G G A C C A	239
QY	249	G C G G C G C C A G C G C A C A G A C A G C G C G C C C C T C T G C C C C G C C T C A A G C G	308
Db	240	G C G G C G C - - A G C G A N C A G A C A G A C A - - G C C G C C C C T C T G C C C C G C C T C A A G C G	296
QY	309	G C G G A C T T C A C C C C C C G A G C T G C G G C G T T C G A C G G C T C C A G A C C C G C G A T A C T	368
Db	297	C G G A N C T T C A C C C C C G A G C T G C G G C G T T C G A C G C G T C C A G G A C C C G C G A T A C T	356
QY	369	C A T G G C C A T C A A C G C A A G T T C G A T G T C A C C A A G G C C G C A A T T C T A C G G G C C G A	428
Db	357	C A T G G C C A T C A A C G C A A G T T C G A T G T C A C C A A G G C C G C A A T T C T A C G G G C C G A	416
QY	429	G G G G C C G T A T G G G G T C T T T G C T G G A A G A G A T C A T C C A G G G C C P T G C C A C A T T T T G C T	488
Db	417	G G C G T A N T - G G G G T C T T T G C T G G A G A G A T C A T C C A G G G C C T T G C A C A A T T T T G C T	475
QY	489	G G A T A A G A A G A C A C T G A A G G A T G A T A G A T G A C C T T T T C T G A C C T C A C T G T G C C	544
Db	476	G G A T A A G A A G A C A C T G A A G G A T G A T A G A T A G A C T T T T C T G A C C T C A C T G T G C C	531
RESULT	3		
LOCUS	A1394728/c		
DEFINITION	A1394728 567 bp mRNA EST 30-MAR-1999 tq24q02.x1 NCI-CGAP CLL1 Homo sapiens cDNA clone IMAGE:2109746 3' similar to TR:000264 000264 PUTATIVE PROGESTERONE BINDING PROTEIN.		
ACCESION	A1394728		
VERSION	A1394728.1 GI:4224275		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
AUTHORS	NCI (bases 1 to 567)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpur Gene Index Unpublished (1997) On Mar 9, 1998 this sequence version replaced gi:2946796. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://bio.lnl.gov/bbr/image/image.html">www.bio.lnl.gov/bbr/image/image.html</a>		
JOURNAL			
COMMENT			

```

Insert Length: 845      Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 438.
Location/Qualifiers
1..567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2109746"
/clone_lib="NCI CGAP_CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1 Not 1; Site 2: Eco RI; 1st strand
CDNA"

```

was primed with a Not I - oligo(dn) primer [5', TGTTACCAATCTGAAGTGGGCGGCGCATGCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 139 a 139 c 123 g 166 t

ORIGIN

Query Match 55.8%; Score 440; DB 45; Length 567;  
Best Local Similarity 97.7%; Pred. No. 7e-92;  
Matches 476; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

```

QY 301 CTCGAAGCGCGGCGACTTCACCCCGCGGAGCTGGCGGCTTCGAGCGGTCGACGAGACCG 360
Db 567 CTCGAAGCGCGGCGACTTCACCCCGCGGAGCTGGCGGCTTCGAGCGGTCGACGAGACCG 508
QY 361 CGCATACTCATGGCCATCAACGCAAGGTGTCGATGTGACAAAGCGCCAAATTCAC 420
Db 507 CGCATACTCATGGCCATCAACGCAAGGTGTCGATGTGACAAAGCGCCAAATTCAC 448
QY 421 GGCCCGGAGGCGCGTATGGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTGCCACA 480
Db 447 GGCCCGGAGGCGCGTATGGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTGCCACA 388
QY 481 TTTTGCTGTGATAGGAGCACTCAAGGATGAGTACGATGACCTTTCGACTCACTGCT 540
Db 387 TTTTGCTGTGATAGGAGCACTCAAGGATGAGTACGATGACCTTTCGACTCACTGCT 328
QY 541 GCCCAGCAGAGACTCTGAGTGACTGGGAGTCTCAGTTCACCTTTCAAGTATCATCACGTG 600
Db 327 GCCCAGCAGAGACTCTGAGTGACTGGGAGTCTCAGTTCACCTTTCAAGTATCATCACGTG 268
QY 601 GGCAAACTGCTGAAGGAGGAGGAGGCCACTGTGTACTCAGATGAGGAAGAACCAAAA 660
Db 267 GGCAAACTGCTGAAGGAGGAGGAGGCCACTGTGTACTCAGATGAGGAAGAACCAAAA 208
QY 661 GATGAGAGTTCGCGGAAAAATG-TTAAAGCAATCAGTGAAGATATATCTA-TNNGTGATT 718
Db 207 GATGAGAGTTCGCGGAAAAATGATTAAGCAATCAGTGAAGATATATCTA-TNNGTGATT 148
QY 719 TTGCAAAATCATTTGTAACAGTCCACTGTGCTTTTAAACATAGTG-TTAAATATTTAG 777
Db 147 TTGCAAAATCATTTGTAACAGTCCACTGTGCTTTTAAACATAGTGATTAATATTTAG 88
QY 778 AAAGTTT 784
Db 87 AAAGTTT 81

```

RESULT 4  
AL035776  
LOCUS 478 bp mRNA EST 27-SEP-1999  
DEFINITION DKF2564M0882\_r1 564 (synonym: hfr2) Homo sapiens cDNA clone  
ACCESSION DKF2564M0882 5', mRNA sequence.  
VERSION AL035776  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 478)  
AUTHORS Poustka A., Klein M., Mewes H.W., Gassenhuber J. and Wiemann S.  
TITLE EST (Poustka, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On Jul 7, 1999 this sequence version replaced gi:5405413.  
Contact: Poustka A.J.  
Department Lehrach  
Max-Planck-Institute for Molecular Genetics  
Ihnestrasse 73, 14195 Berlin, Germany

Tel: +49-30-84131623  
Fax: +49-30-84131128

Email: poustka@mpg-berlin-dahlem.mpg.de  
This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ within the cDNA sequencing consortium of the German Genome Project.

s1 sequence also available.

This clone is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

FEATURES

source

1. 478

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKF2564M0882"

/clone\_lib="564 (synonym: hfr2)"

/tissue\_type="brain"

/dev\_stage="fetal"

/lab\_host="X1-2blue"

/note="Vector: pAMP1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 79 a 157 c 149 g 85 t 8 others

ORIGIN

Query Match 55.5%; Score 437; DB 64; Length 478;

Best Local Similarity 93.9%; Pred. No. 3.3e-91;

Matches 449; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 14 GCGGCGCACTCGCTCGCTCAGAGGAGGAGAAAGTGGCGAGTCCGATCCCTGCCTAGC 73

Db 1 GGGGCCACTCGCTCGCTCNGAGGAGGAGAAAGTGGNNATTCGGGATCCCTGCCTAGC 60

QY 74 GCGGCCCAACCTTACTCCAGAGATCATGCTGCGGAGGATGTGGTGGCGACTGGCGCG 133

Db 61 GCGGCCCAAGCCTTTACTCCAGAAATCATGCTGCGGAGGTTGGTGGCGACTGGCNC 120

QY 134 ACCCAAGCGATCTGGAGAGCGCGGGCTGCTGCATGAGATTTTACGTCGCGGCTCAACC 193

Db 121 ACCCAAGCGATCTGGAGAGCGCGGGCTGCTGCATGAGATTTTACGTCGCGGCTCAACC 180

QY 194 TGCTGCTCTTGGCTCTGCACTCTTCCGCTCTTCAAGATCGTGGCGGGGACAGCCGG 253

Db 181 TGCTGCTCTTGGCTCTGCACTCTTCCGCTCTTCAAGATCGTGGCGGGGACAGCCGG 240

QY 254 GCGCCAGCGGACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 313

Db 241 GCGCCAGCGGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 300

QY 314 ACTTCACCCCGCGAGCTGCGGCGCTTCGAGCGGCTCCAGAGCCCGCGCATCTATGG 373

Db 301 ACTTCACCCCGCGAGCTGCGGCGCTTCGAGCGGCTCCAGAGCCCGCGCATCTATGG 360

QY 374 CCATCAACGGCAAGGTGTTTCGATGTGACCAAGCGCGCAAAATTCACGGCCCGAGGGG 433

Db 361 CCATCAACGGCAAGGTGTTTCGATGTGACCAAGCGCGCAAAATTCACGGCCCGAGGGG 420

QY 434 CGTATGGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTCCACATTTTGCCTGGA 491

Db 421 CGTATGGGGTCTTTGCTGGAAGAGATGCATCCAGGGGCTTCCACATTTTGCCTGGA 478

RESULT 5

W56474

LOCUS

DEFINITION W56474 520 bp mRNA EST 03-JUN-1996

IMAGE:326569 5', mRNA sequence.

ACCESSION W56474

VERSION W56474.1

KEYWORDS GI:1358332

SOURCE human.

ORGANISM Homo sapiens







Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
COMMENT  
On Sep 12, 1996 this sequence version replaced gl:1406900.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estw@wustl.edu

WARNING: there is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 398.

#### FEATURES

source  
1..501  
/organism="Homo sapiens"  
/db\_xref="GDB:3926826"  
/db\_xref="taxon:9606"  
/clone="IMAGE:548310"  
/clone\_lib="Stratagene neuroepithelium NT2RAM1 937234"  
/dev\_stage="Ntera-2/RAM1 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2 (Ntera-2/cl.D1) precursor cells induced with Retinoic Acid for 1 week, followed by 3 weeks in mitotic inhibitors (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3' 7 others  
96 a 155 c 151 g 92 t

Query Match 52.2%; Score 411.6; DB 28; Length 501;  
Best Local Similarity 95.5%; Pred. No. 2.5e-85;  
Matches 484; Conservative 0; Mismatches 16; Indels 7; Gaps 6;

QY 48 TGCCGAGTTCGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGCTGC 107  
Db 1 TGCCGAGTTCGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGCTGC 60  
QY 108 CGAGGATGTGGTGGGACTGGCGCGGACCCAGCGATCTGAGAGCGGGGGCTGTGCA 167  
Db 61 CGAGGATGTGGTGGGACTGGCGCGGACCCAGCGATCTGNAGAGCGGGGGCTGTGCA 120  
QY 168 TGAGATTTTACATCTCCCGCTCAACCTGCTGCTGGCTCTGCTATCTCTGCTCTA 227  
Db 121 TGAGATTTTACATCTCCCGCTCAACCTGCTGCTGGCTCTGCTATCTCTGCTCTA 180  
QY 228 CAAGATCTGCGCGGGGACCGAGCGCGGCGGACGAGGACGACGACGANGCGGCC 287  
Db 181 CAAGATCTGCGCGGGGACCGAGCGCGGCGGACGAGGACGACGACGANGCGGCC 238  
QY 288 CCCTCTGCCCGCCCTCAAGCGGGCGGACTTACCCCGCGAGCTGGCGGCTTCGACGG 347  
Db 239 CCCTCTGCCCGCCCTCAAGCGGGCGGACTTACCCCGCGAGCTGGCGGCTTCGACGG 297  
QY 348 CGTCCAGGACCCCGGCATCTATGCGCATCAACGGCAAGGTGTTCTGATGACCAAGG 407  
Db 298 CGTCCAGGACCCCGGCATCTATGCGCATCAACGGCAAGGTGTTCTGATGACCAAGG 357  
QY 408 CCGCAATTTCTACGGGCGCCGAGGGCGGTATGGGTCTTTTCTGTTGAAGAGATGATCCAG 467  
Db 358 CCGCAATTTCTACGGGCGCCGAGGGCGGTATGGGTCTTTTCTGTTGAAGAGATGATCCAG 415

QY 468 GGCCTTGCCACATTTTTCCTGGATAGGAGCACTGAAGGATGAGT-ACGATGACCTTT 526  
Db 416 GGGCTTG-CACATTTTTCCTGGATAGGAGCACTGAAGGATGAGT-ACGATGACCTTT 474  
QY 527 CTGACCTACTGCTGCCAGCAGGAGA 553  
Db 475 CTGACCTAATGCTGCCAGCAGGAGA 501

#### RESULT

AI885620/c 9  
LOCUS  
DEFINITION  
w159609.x1 NCI-CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2429224 3'  
similar to TR:000264 O00264 PUTATIVE PROGESTERONE BINDING PROTEIN.  
; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AI885620 573 bp mRNA EST 01-SEP-1999  
AI885620.1 GI:5590784  
EST.  
Homo sapiens

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 573)

#### REFERENCE

AUTHORS  
TITLE  
NATIONAL CANCER INSTITUTE / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)

#### JOURNAL

COMMENT  
On May 18, 1998 this sequence version replaced gi:3136735.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonalido, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 368.

#### FEATURES

source

1..573  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2429224"  
/clone\_lib="NCI-CGAP\_Brn25"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCACTCAAGTGGAGCGGCGCATAGTTTGTGTGTGTGTGTGTGTGTGTGT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonalido."

BASE COUNT 144 a 137 c 130 g 162 t  
ORIGIN

Query Match 51.7%; Score 407.6; DB 62; Length 573;  
Best Local Similarity 94.6%; Pred. No. 2.2e-84;  
Matches 474; Conservative 0; Mismatches 22; Indels 5; Gaps 5;

QY 287 CCCCTCTGCCCGCTCAAGCGGCGGACTTACCCCGCGAGCTGGCGGCTTCGACG 346  
Db 569 CCCCTCTGCCCGCTCAAGCGGCGGACTTACCCCGCGGACTTACCCCGCGGCTTCG 511

QY 347 GCGTCCAGGACCCCGCATCTATCGCCATCAACGGCAAGGTGTGATGTGACCAAAG 406  
|||||  
Db 510 GCGTCCA-GACCCGCGCATCTATCGCCATCACTGGCAAGGTGTGATGTGACCAAAG 452  
QY 407 CCGCAAAATCTACGGCCCGAGGCGCGTATGGGTCTTTGCTGGAAGAGATGATCCA 466  
|||||  
Db 451 CCGCAAAATATACGGCCCGAGGCGCGTATGGGTCTTTGCTGGAAGAGATGATCCA 392  
QY 467 GGGCCCTTGCACATTTTTCCTGTGATAGGAGCAGTCAAGGATGAGTACGATGACCTTT 526  
|||||  
Db 391 GGGCCCTTGCACATTTTTCCTGTGATAGGAGCAGTCAAGGATGAGTACGATGACCTTT 332  
QY 527 CTGACCTACTGTGCGCCAGCAGGAGACTCTGAGTACTGGAGTCTCAGTTCACCTTCA 586  
Db 331 CTGACCTACTGTGCGCCAGCAGGAGACTCTGAGTACTGGAGTCTCAGTTCACCTTCA 272  
QY 587 AGTATCATCAGTGGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTACTCAGATG 646  
Db 271 AGTATCATCAGTGGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTACTCAGATG 212  
QY 647 AGGAAGAACCAAAAGATGAGAGTTCCCGGAAAATG-TTAAAGCATTCAGTGGAGTATA 705  
Db 211 AGGAAGAACCAAAAGATGAGAGTCCCGGAAAATGATTTAAAGCATTCAGTGGAGTATA 152  
QY 706 TCTA-TNNTGTATTTGCAAAATCATTTGTAAACAGTCCACTNTGTCTTTAAACATAGTG 764  
Db 151 TCTATTTTGTATTTGCAAAATCATTTGTAAACAGTCCACTCTGTCTTTAAACATAGTG 92  
QY 765 -TTACATATTTAGAAAGTTT 784  
Db 91 ATTACAATATTTAGAAAGTTT 71

RESULT 10  
AA023630 501 bp mRNA EST 21-JAN-1997  
LOCUS  
DEFINITION  
clone IMAGE:457281 5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Nov 29, 1993 this sequence version replaced gi:636945.  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:274169  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 494.  
Location/Qualifiers  
1..501  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:457281"  
/clone\_lib="Soares mouse placenta 4NDMP13.5 14.5"

FEATURES  
Source

/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="organ: placenta; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5',  
TGTTACCAATCTGAAGTGGGAGCGCGCGCAAAATTTTCTTTTCTTTTCTTTTCTTTT  
T-3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M.Fatima Bonaldo."  
BASE COUNT 95 a 159 c 152 g 94 t 1 others  
ORIGIN

Query Match 50.5%; Score 398; DB 27; Length 501;  
Best Local Similarity 88.3%; Pred. No. 3.4e-82;  
Matches 431; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 43 GAAAGTGGCGAGTTCGGATCCCTGCTAGCGCGGCCCAACCTTTTACTCCAGAGATCATG 102  
Db 14 GGAAGCGGACTGTTCCGGAGCTCTGCTAGCGCGGCCCAACCTTTTGTCTCCAGAGATCATG 73  
QY 103 GCTCCGAGGATGTTGGTGGGACTGGCGCCGACCCCAAGGATCTGGAGAGCGGCGGCTG 162  
Db 74 GCTCCGAGGATGTTGGTGGGACTGGCGCCGACCCGAGGAGCTAGAGGGCGGCGGCTG 133  
QY 163 CTGATGAGATTTTCACTGCTCGCGCTCAACCTGCTGCTGGCTCTGATCTTCTCTG 222  
Db 134 CTGACGAGATTTTCACTGCTCTCTCAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 193  
QY 223 CTCTACAAGATCGTGGCGGACCGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 282  
Db 194 CTCTACAAGATCGTGGCGGACCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 253  
QY 283 CGGCCCCCTCTGCGCGGCTCAAGGGGCGGACTTCAACCCCGCGGCGGCGGCGGCGGCTT 342  
Db 254 CCACCCCGCTGCGCGGCTCAAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 313  
QY 343 GAGCGCTCCAGGACCGCGCATCTCATGCGCATCAACGGCAAGGTGTTCGATGTGACC 402  
Db 314 GATGCGTCCAGGACCGCGCATCTCATGCGCATCAACGGCAAGGTGTTCGACGTGACC 373  
QY 403 AAAGGCCCAATTTCTAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 462  
Db 374 AAAGGCCCAAGTTCTAGCGGCGGCTGAGGGCCATATGGGCTTTTCCGGAAGAGATGCA 433  
QY 463 TCCAGGGCGCTTGGCAGATTTTGGCTGGATAGGAAGCACTGAAGGATGAGTACGATGAC 522  
Db 434 TCCAGGGCGCTTGGCAGATTTTGGCTGGCAAAAGCACTGAAGGATGAGTATGACGAC 493  
QY 523 CTTTCTGA 530  
Db 494 CTTTCTGA 501

RESULT 11  
AA288129 508 bp mRNA EST 11-APR-1997  
LOCUS  
DEFINITION  
vbl2e11.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:748748 5',  
similar to TR:E247050 E247050 CHROMOSOME XVI READING FRAME ORF  
YPL170W. ;, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

**AUTHORS** Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,P., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

**TITLE** The WashU-HMMI Mouse EST Project  
**JOURNAL** Unpublished (1996)  
**COMMENT** On Sep 12, 1996 this sequence version replaced gi:1397407.  
Contact: Maïra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MG1-457732

**Seq primer:** -28ml3 rev2 ET from Amersham  
High quality sequence stop: 465.

**FEATURES**

source	Location/Qualifiers
	1..508
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="IMAGE:748748"
	/clone_lib="Soares mouse NML"
	/tissue_type="Liver"
	/lab_host="DH10B"
	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dn) primer [5' TGTTACCAACTCTGAAGTGAGCGGGCCGGAATCTTTTTTTTTTTT 3'] ; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

**BASE COUNT** 97 a 163 c 154 g 94 t

**ORIGIN**

Query Match 48.6%; Score 383; DB 31; Length 508;  
Best Local Similarity 89.38; Pred.No.1e-78;  
Matches 457; Conservative 0; Mismatches 51; Indels 4; Gaps 4;

```

QY 37 GGAGGAGAAAGTCGCAGATTCCCGAGTCCCTGCCCTAGCGCGGCCAACCACTTTACTCCAGAG 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGGAGGAAGCGGACTGTTCGCGAGCTCTGCCTAGCGGGSCCAACTTTTGTCTCCAGAG 60

QY 97 ATCATGGCTGCCGAGGATGTGGTGCGACTGGCGCGGACCACCAACCGATCTGGAGAGGGC 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATCATGGCTCCCGAGGATGTGTGCGGACTGGCGCGGACCACCGGAGGAGCTAGAGGGGGC 120

QY 157 GGGCTGCTGATGAGATTTTCACGTCGCGCGCTCAACTGCTGCTGTGGCTCTGCGATC 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCTGCTGACGAGATTTTCACGTCCTCTCAACTGCTCTCTCTGGGCTCTGCGATC 180

QY 217 TTCTGCTCTACAAGATCGTGC CGGGGACACAGCCGGCGGCGACGCGAGCACGACGACG 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 TTCCTGCTCTACAAGATCGTTTCGCGGGGACCAGCCCGGTGCCAGCTGGGACAAACGACG 240

QY 277 ACGANGCGCCCCCTCTGCCCCGCTCAAGGGCGCGGACTTCACCCCGCGGAGCTGGG 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ACGAA -CAACCCC CGCTGCCCCGCGCTCAAGC -GGCGGACTTCACCCCTGGCGAGCTGAG 298

QY 337 CGCTTCGACGCGCTCCAGGACCCCGCATACTCATGGCCATCAACGGCAAGGTCTTCGAT 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CGTTTCGATGGCTCGAGGACCCGCGATCTCATGCCATCAACGGCAAGGTCTTCGAC 358

QY 397 GTGACCAAAGCCCAAAATTCACGGGCCGAGGGGCCGTAATGGGGTCTTTGCTGGAAGA 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GTGACCAAAGCCCAAGTTTCTACGGCTCTAGGGCTGAGGG -CATATGGGCTCTTTGCCGGA 417

```

Qy	457	GATCATCCAGGAGGCCCTTGCCACATTTTGGCTGGATAAGGAAGCACTGAAGGTAGTAC	516
Dd	418	GATCATCCAGGAGGCCCTTG-CACATTTTGGCTGCACAAAGAAGCACTGAAGGTAGTAT	476
Qy	517	GATGACCTTTCTGACCTCACTGCTGCCCCAGCA	548
Dd	477	GACGACCTTTCTGACCTCACCCCCTGCACAGCA	508
RESULT 12			
AA274818			
LOCUS	AA274818	528 bp	EST
DEFINITION	V020212.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:747743 5'		31-MAR-1997
	similar to TR:E247050 E247050 CHROMOSOME XVI READING FRAME ORF		
	YPL170W. ; mRNA sequence.		
ACCESSION	AA274818		
VERSION	AA274818	GI:1914920	
KEYWORDS	EST.		
SOURCE	Mus musculus		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 528)		
	Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,		
	Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,		
	Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,		
	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and		
	Waterston, R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Sep 7, 1995 this sequence version replaced gi:1027400. Contact: Maria M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicinep 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:456727 Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 472.		
FEATURES			
source	1..528		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:747743"		
	/clone_lib="Soares mouse NML"		
	/tissue_type="Liver"		
	/lab_host="DH10B"		
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; left strand cdna was primed with a Not I - oligo(dT) primer [5' TGTCACCAATCGAATGGCGGCGGCGGAATCTTTTTTTTTTTT 3']; double-stranded cdna was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	99 a 169 c 164 g 96 t		
ORIGIN			
Query Match	48.4%;	Score 381.4;	DB 30; Length 528;
Best Local Similarity	89.1%;	Pred. No. 2.4e-78;	
Matches 456; Conservative	0; Mismatches 52;	Indels 4;	Gaps 4;
Qy	34	GAGGAGGAGAAATGCGGAGTTCCTCGGATCCCTAGCGCGGCCCAACCTTTACTCCA	93
Dd	21	GAGGCGGAGAACGGGACTGTGTCGGAGCTCTGCTTAGCGGGCCCAACCTTTGCTCCA	80
Qy	94	GAGATCATGGCTGCGGAGGATGTGGTGGCAGTGGCGCCGACCAACCGATCTGGAGGC	153

```

|||||
Db 81 GAGATCATGCTGCCGAGGAGTGGTGGCGACTGGCGCCGACCGAGCGAGCTAGAGGC 140
QY 154 GCGGGCTGCTGATGAGATTTTACCTCGCCGCTCGCGCTACCTGCTGCTTGGCTCTGC 213
Db 141 GCGGGCTGCTGACGAGATTTTACCTGCTCTCTCTCACTGCTGCTTGGCTCTGC 200
QY 214 ATCTTCTGCTCTACAAAGATCGTGGCGGGGACGAGCGCGGCGGCGAGCGAGGACG 273
Db 201 ATCTTCTGCTCTACAAAGATCGTGGCGGGGACGAGCGCGGCTGCGAGTGGGACAGG 260
QY 274 ACAGGANGCGCGCCCTCTGCGCCCGCTCAAGCGGCGGCGACTTACCGCCCGCGAGCTG 333
Db 261 ACAGCGAA-CCAGCCCGCTGCGCGCGCTCAAGC-GCGCGACTTCAAGCGCGGAGCTG 318
QY 334 CGCGGCTTCCAGCGGCTCCAGGACCGCGGATCATCTGCGCATCAACGGCAAGGTGTC 393
Db 319 AGCGGTTTCATGCGGCTCCAGGACCGCGCATTTCTCATGCGCATCAACGGCAAGGTGTC 378
QY 394 GATGTGACCAAGCGCGCAATTTCTAGCGGCGCGAGGCGGCTATGGGCTTTGCTGGA 453
Db 379 GAGTGACCAAGCGCGCAAGTCTAGCGGCGCGGCTAGGGG-CATATGGGCTTTGCGGGA 437
QY 454 AGAGATGATCCAGGCGCTTGCCACATTTTGGCTGGATGAAGGACACTGAAGGATGAG 513
Db 438 AGAGATGATCCAGGCGCTTG-CACATTTTGGCTGGACAAAGGACACTGAAGGATGAG 496
QY 514 TAGATGATGACCTTTCTGACCTCACTGCTGCCA 545
Db 497 TATGACGACCTTTCTGACCTCACTGCTGCCA 528

RESULT 13
AA117455
LOCUS
DEFINITION
  mn28g11.r1 Beddington mouse embryonic region Mus musculus CDNA
  clone IMAGE:539300 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING
  ATPASE ;, mRNA sequence.
ACCESSION
  AA117455
VERSION
  AA117455.1 GI:1672468
KEYWORDS
  EST.
SOURCE
  house mouse.
  Mus musculus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
  Waterston, R.
  The WashU-HMI Mouse EST Project
  Unpublished (1996)
  On Sep 12, 1996 this sequence version replaced gi:1393715.
  Contact: Marra M/Mouse EST Project
  WashU-HMI Mouse EST Project
  Washington University School of MedicineP
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:326236
  Seq primer: -40ml3 ET
  High quality sequence stop: 466.
  Location/Qualifiers
    1..512
      /organism="Mus musculus"
      /strain="C57BL6 x DBA"
      /db_xref="taxon:10090"
      /clone="IMAGE:539300"
      /clone_lib="Beddington mouse embryonic region"
FEATURES
  source
    /sex="pooled"
    /tissue_type="embryo"
    /dev_stage="7.5dpc"
    /lab_host="DH12S"
    /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
    SalI; Site_2: NotI; Cloned unidirectionally. Primer:
    Oligo dt. Gastrulating embryos were collected at 7.5dpc
    from C57BL6 x DBA matings, excluding embryos that had
    developed head folds and all extraembryonic tissues.
    Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
    Referenced in Development 121, 2479-2489 (1995)."
    BASE COUNT 134 a 128 c 132 g 117 t 1 others
    ORIGIN
      Query Match 48.3%; Score 380.8; DB 28; Length 512;
      Best Local Similarity 88.3%; Pred. No. 3.3e-78;
      Matches 445; Conservative 0; Mismatches 56; Indels 3; Gaps 3;
      QY 283 CCGCCCTCTGCCCCCGCTCAAGCGCGGCTTACCCCGCGAGCTCGGCGCTTC 342
      Db 6 CCACCCCGCTGCCCCCGCTCAAGC-GGCANCTTACCCCTGCCGAGCTGAGCGGTTTC 64
      QY 343 GACGCGCTCCAGGACCGCGCATATCTATGCCCATCAACGGCAAGGTGTTTCGATGTGACC 402
      Db 65 GATGCGCTCCAGGACCGCGCATTCTCATGCCATCAACGGCAAGGTGTTTCGACGTGACC 124
      QY 403 AAAGCGCGCAAAATTTCTAGCGGCGCGAGGCGGCTATGGGCTTTTCTGGGAAGATGCA 462
      Db 125 AAAGCGCGCAAGTTCTAGCGGCGCTAGGGGCTATGGGCTTTTGGCGGAAGATGCA 184
      QY 463 TCCAGGGCGCTTGCACATTTTGCCTGATGAAGGAAGCACTGAAGGATGACTAGCATGAC 522
      Db 185 TCCAGGGCGCTTGCACATTTTGCCTGATGAAGGAAGCACTGAAGGATGACTAGCATGAC 244
      QY 523 CTTTCTGACCTCACTGCTGCGGCGGAGAGACTCTGAGTGAAGGAGGAGGAGGAGGAGGAGG 582
      Db 245 CTTTCTGACCTCACTGCTGCGGCGGAGAGACTCTGAGTGAAGGAGGAGGAGGAGGAGG 304
      QY 583 TTCAAGTATCATCAGTGGGCAAACTGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 642
      Db 305 TTCAAGTATCATCAGTGGGCAAACTGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 364
      QY 543 GATGAGGAAGCAAAAGATGAGAGTCCCGGAAATG-TTAAAGCATTCAGTGGAG 701
      Db 365 GATGATGAAGCAAAAGATGAGAGTCCCGGAAATGATGAATGAAGCATTCGCTGGAGC 424
      QY 702 TATATCATNTGATTTTGCATAATCATTTGTAAACAGTCCACTGTGCTTTAAACACATA 761
      Db 425 ATATCTATTTGATTTTGCATAATCATTTGTAAACAGTCCACTGTGCTTTAAACACATG 484
      QY 762 GTG-TTACAATATTTAGAAAGTTT 784
      Db 485 GTGATTTCAATATTTAGAAAGTTT 508

RESULT 14
AA101294
LOCUS
DEFINITION
  zn71f03.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
  cDNA clone IMAGE:563645 5' similar to WP:K07E3.6 CE04722
  TRANSLOCATING ATPASE ;, mRNA sequence.
ACCESSION
  AA101294
VERSION
  AA101294.1 GI:1648032
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
  Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
  Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

```

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

On May 18, 1995 this sequence version replaced gi:811508.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 2058 Std Error: 0.00

Seq primer: -28M13 rev2 from Amerham.

#### FEATURES

Location/Qualifiers

1..429

/organism="Homo sapiens"

/db\_xref="GDB:4596168"

/db\_xref="taxon:9606"

/clone\_image="563645"

/clone\_lib="Stratagene NT2 neuronal precursor 937230"

/tissue\_type="neuroepithelial cells"

/dev\_stage="Ntera-2 neuroepithelial cells"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: brain; Vector: pBluescript SK-; Site: 1;

EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Uninduced, exponentially growing neuroepithelial

cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG

3' -3' adaptor sequence: 5' CTCAGATTTTTTTTTTTT 3'

79 a 136 c 131 g 9 others

BASE COUNT

ORIGIN

Query Match 47.3%; Score 373; DB 28; Length 429;  
Best Local Similarity 94.0%; Pred. No. 2e-76;  
Matches 404; Conservative 0; Mismatches 24; Indels 2; Gaps 2;

QY 42 AGAAAGTGGCGAGTCCCGATCCCTGCTAGCGCGGCCCAACCTTACTCCAGATCAT 101  
Db 1 AGAAAGTGGCGAGTCCCGATCCCTGCTAGCGCGGCCCAACCTTACTCCAGATCAT 60

QY 102 GGTCTCGAGGATGTGGTGGCGACTGGCGCGGACCAAGCGATCTGGAGCGCGGGCT 161  
Db 61 GGTCTCGAGGATGTGGTGGCGACTGGCGCGGACCAAGCGATCTGGAGCGCGGGCT 120

QY 162 GGTCTCGAGGATTTTACGTGGCGGCTCAACCTGCTGCTTGGGCTC-TGCATCTTC 220  
Db 121 GGTCTCGAGGATTTTACGTGGCGGCTCAACCTGCTGCTTGGGCTC-TGCATCTTC 180

QY 221 TGTCTTACAGATCGTGGCGGGGACGACCGCGCGGACCGGCGGACGACGACGACGA 280  
Db 181 TGTCTTACAGATCGTGGCGGGGACGACCGCGCGGACCGGCGGACGACGACGACGA 240

QY 281 NGCGCGCCCTCTGCGCGGCTCAAGCGGCGGACTTACCCCGCGGAGCTGCGGGCT 340  
Db 241 -GCGCGCCCTCTGCGCGGCTCAAGCGGCGGACGACGCGGAGCTGCGGGCT 299

QY 341 TCGACGCGCTCCAGGACCGCGGATCTCATGCGGATCAACGCGCAAGGTTTCGATGTGA 400  
Db 300 TCGACGCGCTCCAGGACCGCGGATCTCATGCGGATCAACGCGCAAGGTTTCGATGTGA 359

QY 401 CCAAGCGCGCAAAATCTACGGGCGCGGAGGCGCGTATGGGGTCTTTGCTGGAGAGATG 460  
Db 360 CCAAGCGCGCAAAATCTACGGGCGCGGAGGCGGTAATTTGGGGTCTTTGCTGGAGAGATG 419

QY 461 CATCCAGGGG 470  
Db 420 CATCCAGGGG 429

#### RESULT 15

N52291/c

LOCUS

DEFINITION

IMAGE:245950 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

N52291 458 bp mRNA EST 28-JAN-1997  
yv48a12.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:245950 3', mRNA sequence.

N52291

N52291.1 GI:1193457

human.

EST.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 458)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chitapelli, B.,

Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

On Sep 21, 1992 this sequence version replaced gi:276109.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 826 Std Error: 0.00

Seq primer: m13 -40 forward

High quality sequence stop: 278.

Location/Qualifiers

1..458

/organism="Homo sapiens"

/db\_xref="GDB:3795196"

/db\_xref="taxon:9606"

/clone\_image="245950"

/clone\_lib="Soares fetal liver spleen INFLS"

/sex="male"

/dev\_stage="20 week-post conception fetus"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)

with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer

[5' AACTGGAAGATTAATTAAGATCTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 102 a 126 c 114 g 114 t 2 others

ORIGIN

Query Match 47.1%; Score 371.4; DB 25; Length 458;  
Best Local Similarity 96.3%; Pred. No. 4.7e-76;  
Matches 442; Conservative 0; Mismatches 10; Indels 7; Gaps 6;

QY 300 CCTCAAGCGCGGACTTCA-CCCCCGCGAGCTCGCGGCTTCGAGGGTCCAGGACC 358

Db 458 CCTCAAGCGCGGACTTCA-CCCCCGCGAGCTCGCGGCTTCGAGGGTCCAGGACC 399

QY 359 C--CGGCATACTCATGGCCATCAACGCAAGGTTTCGATGTGACCAAGGCCCA-AAT 415

Db 398 CCGGGCCATCTCATGGCCATCAACGCAAGGTTTCGATGTGACCAAGGCCCAAT 339

QY 416 TCTACGGCGCGGCGCGCTATGGGCTCTTTGTTGAAGAGATGTCATCCAGGGGCTTG 475

Db 338 TCTACCGGCGNGAGGGCCGCTATGGGCTCTTGGTGAAGAGATGCATCCA-GGGCCCTG 280  
QY 476 CCACATTTTGCCTGGATAAGGAAGCACTGAAGGATGAGTAGCATGACCTTTCTGACCTCA 535  
Db 279 CCACATTTTGCCTGGATAAGGAAGCACTGAAGGATGAGTAGCATGACCTTTCTGACCTCA 220  
QY 536 CTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTCAAGTATCATC 595  
Db 219 CTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTCAAGTATCATC 160  
QY 596 ACSTGGCCAACTCCTCTGAAGAGGGGGAGGAGCCCACTGTCTACTCAGATGAGGAAGAAC 655  
Db 159 ACSTGGCCAACTCCTCTGAAGAGGGGGAGGAGCCCACTGTCTACTCAGATGAGGAAGAAC 100  
QY 656 CAAAAGATGAGAGTTCCTGGGAAAAATG-TTAAAGCATTCACTGGAAGTATATCTA-TNNT 713  
Db 99 CAAAAGATGAGAGTTCCTGGGAAAAATGATTAAAGCATTCACTGGAAGTATATCTATTTT 40  
QY 714 GTATTTTGAAAAATCATTGTGAACAGTCCCACTNTGTCTT 752  
Db 39 GTATTTTGAAAAATCATTGTGAACAGTCCCACTCTGTCTT 1

Search completed: April 19, 2000, 01:52:18  
Job time: 3045 sec

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1121	99.3	220	1	W71362	Human cytokine/ste
2	983	87.1	195	1	W85000	Human stomach carc
3	976	86.4	195	1	W80396	A secreted protein
4	943.5	83.6	194	1	W39900	Pig plasma membran
5	943.5	83.6	194	1	W54363	Steroid membrane b
6	501	44.4	170	1	W67874	Human secreted pro
7	249	22.1	52	1	W39901	Pig plasma membran
8	102	9.0	19	1	W39902	Pig plasma membran
9	88.5	7.8	176	1	W84621	Ehrlichia sp. HGE-
10	85.5	7.6	325	1	R5028	Protein with Oxa2a
11	79	7.0	1073	1	P60569	Sequence of the R1
12	77	6.8	334	1	R80220	GRB-7 adaptor prot
13	77	6.8	335	1	R80167	Mouse signal trans
14	77	6.8	335	1	R80161	GRB-7 central BLM
15	77	6.8	334	1	R80164	Mouse signal trans
16	77	6.8	535	1	R86900	Human GRB-7. DNA e
17	76	6.7	502	1	W80398	A secreted protein
18	74	6.6	249	1	P91215	Domain VI of mouse
19	74	6.6	1607	1	W50897	Mouse laminin G1 c
20	73.5	6.5	448	1	W53463	Human gp49 Hm18 po
21	72.5	6.4	448	1	W82551	Human LiR-pbm2 pro
22	72.5	6.4	873	1	W49031	Human Prt1-like su
23	71.5	6.3	478	1	W92950	WO9905287 Seq ID 1
24	71.5	6.3	1231	1	Y00219	Enterococcus faeca
25	71.5	6.3	1265	1	Y00218	Enterococcus faeca
26	71.5	6.3	1312	1	R53787	Deep Vent DNA poly
27	71	6.3	166	1	P50030	Sequence of new mo
28	71	6.3	709	1	R38744	Human SSRP. Ident
29	71	6.3	709	1	W39212	Human SSRP1 protei
30	70.5	6.2	1609	1	W50898	Human laminin G1 c
31	70	6.2	848	1	W82600	Starch branching e
32	69.5	6.2	446	1	W85119	A delta-5 desatur
33	69.5	6.2	446	1	W95506	Mortierella alpina
34	69	6.1	155	1	W20632	H. pylori derived

Db 121 ASRGLATFCLDKALKDEYDDLSDLTAAQOETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 QY 181 SDEEPPKDESSRKNKAFSGSISXXYFAKSFVTVHXVFKT 220  
 |||||||  
 Db 181 SDEEPPKDESSRKNKAFSGSISXXYFAKSFVTVHXVFKT 220  
 |||||||

RESULT 2  
 W88500  
 ID W88500 standard; Protein; 195 AA.  
 AC W88500;  
 DT 30-MAR-1999 (first entry)  
 DE Human stomach carcinoma clone HP10413-encoded protein.  
 KW Transmembrane protein; HP10413; human; stomach cancer.  
 OS Homo sapiens.  
 PN W0985508-A2.  
 PD 10-DEC-1998.  
 PF 03-JUN-1998; J02445.  
 PR 03-JUN-1997; JP-144948.  
 PA (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENTRE.  
 PI Kato S, Sekine S, Yamaguchi T;  
 DR WPI: 99-045730/04.  
 DR N-PSDB; W84368.  
 CC New human proteins containing transmembrane domains and their  
 PT encoding sequences - useful in the preparation of antibodies and  
 PT large-scale protein production, gene diagnosis, and gene therapy  
 PS Claim 1: Page 138-139; 178pp; English.  
 CC This is the amino acid sequence of a transmembrane protein encoded  
 CC by human stomach cancer cDNA clone HP10413 (see W84368). The  
 CC encoded protein (28 kDa) has one transmembrane domain at the  
 CC N-terminal. It shows 96.4% homology to the pig steroidal  
 CC membrane-binding protein. The invention provides nucleotide  
 CC sequences (see W84359-76) coding for 18 transmembrane proteins  
 CC (see W88491-508), vectors containing such polynucleotides, and  
 CC eukaryotic cells containing the vectors. The proteins can be  
 CC used as antigens or as compositions in the preparation of  
 CC antibodies against the proteins. The polynucleotides can be used  
 CC as probes for gene diagnosis, and as gene sources for gene therapy  
 CC and large-scale production of proteins encoded by the cDNA. The  
 CC host cells are used for the detection of ligands corresponding to  
 CC the expressed proteins, and the screening of low mol.wt. medicines.  
 SQ Sequence 195 AA;

Query Match 87.1%; Score 983; DB 1; Length 195;  
 Best Local Similarity 96.9%; Pred. No. 1.8e-103;  
 Matches 188; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAEDVYVATGADPSDLESGLLHEIFTSPNLNLLGLCIFYLYKIVRGDQPAASGDRTTT 60  
 |||||||  
 Db 1 MAEDVYVATGADPSDLESGLLHEIFTSPNLNLLGLCIFYLYKIVRGDQPAASGSDDD 60  
 |||||||

QY 61 XPPPLPKLRKRDFTPAELRRFDGVQDPRILMAINGKVFDTVKGRFYGPEGYGVFAGRD 120  
 |||||||  
 Db 61 EPPPLPKLRKRDFTPAELRRFDGVQDPRILMAINGKVFDTVKGRFYGPEGYGVFAGRD 120  
 |||||||

QY 121 ASRGLATFCLDKALKDEYDDLSDLTAAQOETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 |||||||  
 Db 121 ASRGLATFCLDKALKDEYDDLSDLTAAQOETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 |||||||

QY 181 SDEEPPKDESSRKN 194  
 |||||||  
 Db 181 SDEEPPKDESSRKN 194  
 |||||||

RESULT 3  
 W80396  
 ID W80396 standard; Protein; 195 AA.  
 AC W80396;  
 DT 12-JAN-1999 (first entry)  
 DE A secreted protein encoded by clone bp646.10.  
 KW Secreted protein; immune stimulating; suppressing;

KW haematopoiesis regulating activity; tissue growth activity; activin;  
 KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;  
 KW thrombolytic activity; anti-inflammatory activity; cadherin;  
 KW tumour invasion suppressor activity; tumour inhibition activity.  
 OS Homo sapiens.  
 FH Key  
 FT Location/Qualifiers  
 FT Misc\_difference 108  
 FT /note= "not defined"  
 PN W09840486-A2.  
 PD 17-SEP-1998.  
 PF 13-MAR-1998; U04977.  
 PR 29-OCT-1997; US-960022.  
 PR 14-MAR-1997; US-815047.  
 PA (GENY) GENETICS INST INC. *102(e) out when issue*  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racine LA, Spaulding V, Treacy M;  
 DR WPI: 98-520812/44.  
 DR N-PSDB; V63170.  
 CC New isolated human poly:nucleotide(s) and secreted proteins -  
 CC obtained from e.g. human foetal kidney, placenta, foetal brain,  
 CC adult testes, adult brain or adult uterus cDNA libraries  
 PS Claim 17; Page 69; 110pp; English.  
 CC The present sequence represents a secreted protein. The nucleic acid  
 CC sequence is isolated from a human foetal kidney cDNA library using  
 CC probe V63180. The polypeptide may have biological activities such as  
 CC e.g. nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity or other activities.  
 SQ Sequence 195 AA;

Query Match 86.4%; Score 976; DB 1; Length 195;  
 Best Local Similarity 96.4%; Pred. No. 1.1e-102;  
 Matches 187; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAEDVYVATGADPSDLESGLLHEIFTSPNLNLLGLCIFYLYKIVRGDQPAASGDRTTT 60  
 |||||||  
 Db 1 MAEDVYVATGADPSDLESGLLHEIFTSPNLNLLGLCIFYLYKIVRGDQPAASGSDDD 60  
 |||||||

QY 61 XPPPLPKLRKRDFTPAELRRFDGVQDPRILMAINGKVFDTVKGRFYGPEGYGVFAGRD 120  
 |||||||  
 Db 61 EPPPLPKLRKRDFTPAELRRFDGVQDPRILMAINGKVFDTVKGRFYGPEGYGVFAGRD 120  
 |||||||

QY 121 ASRGLATFCLDKALKDEYDDLSDLTAAQOETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 |||||||  
 Db 121 ASRGLATFCLDKALKDEYDDLSDLTAAQOETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 |||||||

QY 181 SDEEPPKDESSRKN 194  
 |||||||  
 Db 181 SDEEPPKDESSRKN 194  
 |||||||

RESULT 4  
 W39900  
 ID W39900 standard; Protein; 194 AA.  
 AC W39900;  
 DT 11-JUN-1998 (first entry)  
 DE Pig plasma membrane bound receptor.  
 KW Pig; plasma membrane bound receptor; steroid receptor;  
 KW opioid receptor; antibody; diagnosis; therapy; tumour cell;  
 KW nerve cell.  
 OS Sus scrofa.  
 PN DE19627237-A1.  
 PD 08-JAN-1998.  
 PF 07-JUL-1996; 027237.  
 PR 07-JUL-1996; DE-027237.  
 PA (WEHL/) WEHLING M.  
 PI Wehling M;  
 DR WPI: 98-064316/07.  
 DR N-PSDB; V09852.



PT Nucleic acid encoding plasma membrane bound steroid or opioid  
 PT receptor - and related vectors, poly:peptide(s) and antibodies, used  
 PT for diagnosis and therapy, particularly of tumour cells targetted by  
 PT ligands of receptor  
 PS Claim 12; Pages 11-13; 18pp; German.  
 CC The present sequence is a pig plasma membrane bound steroid or  
 CC opioid receptor, which can be used to generate antibodies for the  
 CC diagnosis and therapy of tumour cells in organs targetted by  
 CC steroid hormones and/or opioids, and nerve cells. Probes and  
 CC primers based on the receptor cDNA can be used to detect expression  
 CC of the receptor, and isolate related coding sequences. It can  
 CC also be used to produce recombinant polypeptides.  
 SQ Sequence 194 AA;

Query Match 83.6%; Score 943.5; DB 1; Length 194;  
 Best Local Similarity 93.3%; Pred. No. 5.1e-99;  
 Matches 181; Conservative 4; Mismatches 8; Indels 1; Gaps 1;  
 QY 1 MAEDVATGADPSDLESGLLHEIFTSPLNLLGLCFLLYKIVRGDQPAASGDRTTT 60  
 DB 1 MAEDVATGADPSELEGGLLHEIFTSPLNLLGLCFLLYKIVRGDQPAAS-DSDDD 59  
 QY 61 XPPLPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPPEYGVFAGRD 120  
 DB 60 EPPPLPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPPEYGVFAGRD 119  
 QY 121 ASRGLATFCLDKALKDEYDDLSLTAQAQETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 DB 120 ASRGLATFCLDKALKDEYDDLSLTAQAQETLSDWESQFTFKYHHVHGKLLKEGEPTVY 179  
 QY 181 SDEEPPKDESSRKN 194  
 DB 180 SDEEPPKDESARKN 193

RESULT 5  
 W54363  
 ID W54363 standard; protein; 194 AA.  
 DT 14-AUG-1998 (first entry)  
 DE Steroid membrane binding protein.  
 KW Endometrium; hyperplasia; adenocarcinoma; proliferative phase;  
 KW 2D gel electrophoresis; detection.  
 OS Homo sapiens.  
 PN W09810291-A1.  
 PD 12-MAR-1998.  
 PF 05-SEP-1997; G02394.  
 PR 08-APR-1997; GB-007132.  
 PR 06-SEP-1996; GB-018600.  
 PA (CLIN-) CENT CLINICAL & BASIC RES.  
 PI Byrjalsen I, Fey SJ, Larsen P;  
 DR WPI; 98-207057/18.  
 PT Biochemical markers of human endometrium - useful for, e.g.  
 PT diagnosis of hyperplasia and adenocarcinoma  
 PS Disclosure; Page 23; 77pp; English.  
 CC Proteins W54363-W54364 are examples of proteins produced in the  
 CC endometrium during the hyperplasia, adenocarcinoma or proliferative  
 CC phase of the endometrium. The presence and quantities of these proteins  
 CC can be detected using 2D gel electrophoresis comparison of cell lysates.  
 CC The proteins can be used as biochemical markers to detect the phase of  
 CC the endometrium and can be measured in body fluids, obviating the need  
 CC for endometrial biopsies.  
 SQ Sequence 194 AA;

Query Match 83.6%; Score 943.5; DB 1; Length 194;  
 Best Local Similarity 93.3%; Pred. No. 5.1e-99;  
 Matches 181; Conservative 4; Mismatches 8; Indels 1; Gaps 1;  
 QY 1 MAEDVATGADPSDLESGLLHEIFTSPLNLLGLCFLLYKIVRGDQPAASGDRTTT 60  
 DB 1 MAEDVATGADPSELEGGLLHEIFTSPLNLLGLCFLLYKIVRGDQPAAS-DSDDD 59

QY 61 XPPLPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPPEYGVFAGRD 120  
 DB 60 EPPPLPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPPEYGVFAGRD 119  
 QY 121 ASRGLATFCLDKALKDEYDDLSLTAQAQETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
 DB 120 ASRGLATFCLDKALKDEYDDLSLTAQAQETLSDWESQFTFKYHHVHGKLLKEGEPTVY 179  
 QY 181 SDEEPPKDESSRKN 194  
 DB 180 SDEEPPKDESARKN 193

RESULT 6  
 W67874  
 ID W67874 standard; Protein; 170 AA.  
 AC W67874;  
 DT 25-MAR-1999 (first entry)  
 DE Human secreted protein encoded by gene 68 clone HLB049.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; gastritis; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09842738-A1.  
 PD 01-OCT-1998.  
 PF 19-MAR-1998; U05311.  
 PR 30-MAY-1997; US-050937.  
 PR 21-MAR-1997; US-041276.  
 PR 21-MAR-1997; US-041277.  
 PR 21-MAR-1997; US-041281.  
 PR 21-MAR-1997; US-042344.  
 PR 30-MAY-1997; US-048069.  
 PR 30-MAY-1997; US-048094.  
 PR 30-MAY-1997; US-048095.  
 PR 30-MAY-1997; US-048096.  
 PR 30-MAY-1997; US-048099.  
 PR 30-MAY-1997; US-048131.  
 PR 30-MAY-1997; US-048135.  
 PR 30-MAY-1997; US-048154.  
 PR 30-MAY-1997; US-048160.  
 PR 30-MAY-1997; US-048186.  
 PR 30-MAY-1997; US-048187.  
 PR 30-MAY-1997; US-048188.  
 PR 30-MAY-1997; US-048350.  
 PR 30-MAY-1997; US-048351.  
 PR 30-MAY-1997; US-048352.  
 PR 30-MAY-1997; US-048355.  
 PR 05-AUG-1997; US-054804.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Brewer JA, Duan R, Ebner R, Ferrie AM, Florence KA,  
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,  
 PI Rosen CA, Ruben SM, Shi Y, Young P;  
 DR WPI; 99-070066/06.  
 N-PSDB; X00678.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 11; Page 308-309; 385pp; English.  
 CC This sequence represents a secreted human protein encoded by the gene  
 CC clone detailed in the descriptor line. The gene can be used to generate  
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
 CC portion (e.g. X00602) for increasing the stability of the fused protein  
 CC as compared to the human protein only.  
 CC The invention relates to 87 novel genes and their fragments (nucleic acid  
 CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be

CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 87 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see X00611 for described  
 CC uses).  
 SQ Sequence 170 AA;

Query Match 44.4%; Score 501; DB 1; Length 170;  
 Best Local Similarity 56.9%; Pred. No. 5.1e-49;  
 Matches 95; Conservative 30; Mismatches 38; Indels 4; Gaps 1;  
 QY 30 LNLALLGICIFLLYXI-----VRGDPAASGDRTTTTXPPPLPKRRDFTPAELRRFDGVQ 85  
 Db 3 LNLVALVLLGAYRLVWRGRRGLGACAGAGEESPATSLPRMKRRDFSLEQLROYDGSR 62  
 QY 86 DPRILMAINGKVFDTGKRKYGEGPGYVAGRADASGLATFCLDKALDEYDLSL 145  
 Db 63 NPIILLAVNGKVFDTGSKFYGPAGPYGIFAGRADASGLATFCLDKALRDEYDLSL 122  
 QY 146 TAAQOETLSDVSEFTFKYHHVGLLKEGEPTVYSDDEEPKDESSR 192  
 Db 123 NAVONESVREWEQFKEKYDIVGRLLKPEEPSEYTDDEETKDHNQ 169

RESULT 7  
 W39901  
 ID W39901 standard; Protein; 52 AA.  
 AC W39901; 1998 (first entry)  
 DE Pig plasma membrane bound receptor fragment (residues 2 to 53).  
 KW Pig; plasma membrane bound receptor; steroid receptor;  
 KW opioïd receptor; antibody; diagnosis; therapy; tumour cell;  
 KW nerve cell.  
 OS Sus scrofa.  
 PN DE19627237-A1.  
 PD 08-JAN-1998.  
 PF 07-JUL-1996; 027237.  
 PR 07-JUL-1996; DE-027237.  
 PA (WEHL/) WEHLING M.  
 PI Wehling M;  
 DR WPI; 98-064316/07.  
 PT Nucleic acid encoding plasma membrane bound steroid or opioïd  
 PT receptor - and related vectors, poly:peptide(s) and antibodies, used  
 PT for diagnosis and therapy, particularly of tumour cells targetted by  
 PT ligands of receptor.  
 PS Claim 1; Page 10; 18pp; German.  
 CC The present sequence is a pig plasma membrane bound steroid or  
 CC opioïd receptor fragment, which can be used to generate antibodies  
 CC for the diagnosis and therapy of tumour cells in organs targetted  
 CC by steroid hormones and/or opioïds, and nerve cells. Probes and  
 CC primers based on the receptor cDNA can be used to detect expression  
 CC of the receptor, and isolate related coding sequences. It can  
 CC also be used to produce recombinant polypeptides.  
 SQ Sequence 52 AA;

Query Match 22.1%; Score 249; DB 1; Length 52;  
 Best Local Similarity 94.2%; Pred. No. 3.1e-21;  
 Matches 49; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAEDVATGADPSLEGGLLHEIFTSPNLNLLGLCIFLLYKIVRGDPA 53  
 Db 1 AAEDVATGADPSLEGGLLHEIFTSPNLNLLGLCIFLLYKIVRGDPA 52  
 RESULT 8  
 W39902  
 ID W39902 standard; Protein; 19 AA.  
 AC W39902;  
 DT 11-JUN-1998 (first entry)  
 DE Pig plasma membrane bound receptor fragment (residues 90 to 108).  
 KW Pig; plasma membrane bound receptor; steroid receptor;

KW opioïd receptor; antibody; diagnosis; therapy; tumour cell;  
 KW nerve cell.  
 OS Sus scrofa.  
 PN DE19627237-A1.  
 PD 08-JAN-1998.  
 PF 07-JUL-1996; 027237.  
 PR 07-JUL-1996; DE-027237.  
 PA (WEHL/) WEHLING M.  
 PI Wehling M;  
 DR WPI; 98-064316/07.  
 PT Nucleic acid encoding plasma membrane bound steroid or opioïd  
 PT receptor - and related vectors, poly:peptide(s) and antibodies, used  
 PT for diagnosis and therapy, particularly of tumour cells targetted by  
 PT ligands of receptor.  
 PS Example 2; Page 5; 18pp; German.  
 CC The present sequence is a pig plasma membrane bound steroid or  
 CC opioïd receptor fragment, which can be used to generate antibodies  
 CC for the diagnosis and therapy of tumour cells in organs targetted  
 CC by steroid hormones and/or opioïds, and nerve cells. Probes and  
 CC primers based on the receptor cDNA can be used to detect expression  
 CC of the receptor, and isolate related coding sequences. It can  
 CC also be used to produce recombinant polypeptides.  
 SQ Sequence 19 AA;

Query Match 9.0%; Score 102; DB 1; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 91 MAINGKVFDTGKRKYGP 109  
 Db 1 MAINGKVFDTGKRKYGP 19  
 RESULT 9  
 W82621  
 ID W82621 standard; Protein; 176 AA.  
 AC W82621;  
 DT 01-MAR-1999 (first entry)  
 DE Ehrlichia sp. HGE-24 protein.  
 KW Antigen; human granulocytic ehrlichiosis; HGE; diagnosis; treatment;  
 KW immunogenic; infection; vaccine; epitope; medicament; rodent; bacteria;  
 KW protective immunity; Lyme disease; babesiosis; detection; HGE-24.  
 OS Ehrlichia sp.  
 PN W09842740-A2.  
 PD 01-OCT-1998.  
 PF 23-MAR-1998; U05695.  
 PR 20-NOV-1997; US-975762.  
 PR 21-MAR-1997; US-821324.  
 PA (CORI-) CORIXA CORP.  
 PI Houghton R, Lodes MJ, Reed SG;  
 DR WPI; 98-609891/51.  
 DR N-PSDB; V69409.  
 PT Poly:peptide(s) comprising immunogenic portion of Ehrlichia antigen  
 PT - and encoding DNA sequences, useful for e.g. diagnosis and  
 PT treatment of Ehrlichia infection, especially human granulocytic  
 PT ehrlichiosis  
 PS Example 1; Page 40-41; 140pp; English.  
 CC This sequence represents an immunogenic portion of Ehrlichia antigen,  
 CC HGE-24. This polypeptide is useful in the treatment of Ehrlichia  
 CC infection, and as a vaccine for the prevention of infection. Such  
 CC vaccines comprise an immunogenic portion of an Ehrlichia antigen  
 CC associated with human granulocytic ehrlichiosis (or a variant) and is  
 CC thus especially useful in the treatment of human granulocytic  
 CC ehrlichiosis (HGE). The polypeptides, antigenic epitopes or DNA  
 CC molecules can be combined with a suitable carrier in pharmaceutical  
 CC compositions. Such compositions and vaccines are useful to manufacture  
 CC medicaments for inducing protective immunity against Ehrlichia infection  
 CC in patients especially against HGE. HGE is caused by a rodent bacterium  
 CC normally transmitted to humans by the same tick which transmits Lyme  
 CC disease and babesiosis. Co-infection with these diseases is thus possible  
 CC and the compositions of the invention may be used in methods to detect at  
 CC least one of Ehrlichia infection, Lyme disease or B. microti infection in

CC patients.  
SQ Sequence 176 AA;

Query Match 7.8%; Score 88.5; DB 1; Length 176;  
Best Local Similarity 27.0%; Pred. No. 0.025;  
Matches 50; Conservative 19; Mismatches 77; Indels 39; Gaps 9;

Qy 39 IFLYKIVRGDPAASGORTTTPPLRLKRRDTPAELRRFDGVQDPRILMAINGKVF 98  
Db 6 VILLAKELAYDVVTGTDKLTAA---LAKTSKDF--VQFARAVGVSHFN----IDGKVC 56  
Qy 99 DVTGRKFYGPYGVFAGRDASRLATFCLDKKALDEYDLSLTAAQOETLSDWES 158  
Db 57 KTLTGHTSNDSVGVGELTGO-ASASETSLCGK-----GKNSSGGGAPEVLRF-- 107  
Qy 159 QFTFYHHVKKLLKEGEE--PTVYSDEEPPKDES-SRKNVKAFFSGISXXYFAKSPVTYH 215  
Db 107 -----VKRSLKGGQWNPTRATESPPTKSETNDNAK-----VAKDLVDLEN 149  
Qy 216 XVFKT 220  
Db 150 PEKKT 154

## RESULT 10

R52028  
ID R52028 standard; Protein; 325 AA.

AC R52028;  
DT 23-NOV-1994 (first entry)  
DE Protein with Oxetanocin-A producing activity.  
KW Oxetanocin-A; antibiotic resistance; Bacillus megaterium.  
OS Bacillus megaterium.  
FH Key Location/Qualifiers  
FT protein 1..325  
FT /note= "Oxetanocin-A producing activity"  
PN J06078776-A.  
PD 22-MAR-1994.  
PF 03-SEP-1992; 258877.  
PR 03-SEP-1992; JP-258877.  
PA (NIPK) NIPPON KAYAKU KK.  
DR WPI: 94-131277/16.  
DR N-PSDB: Q61885.  
PT Gene involved in oxetanocin-A prodn. - also DNA fragments for resistance to the antibiotic, recombinant vectors and transformed hosts, esp. Bacillus megaterium  
PS Claim 4; Page 14; 18pp; Japanese.  
CC R52028 shows the product of ORF6 of Q61885 (which contains the gene involved in oxetanocin-A production). This protein is encoded on the complementary strand of Q61885. The plasmid containing the sequence was isolated from oxetanocin-A producing Bacillus megaterium NK84-0128. Oxetanocin-A producing microorganisms or oxetanocin-A resistant microorganisms can be detected using the new nucleotide sequence to design probes.  
CC sequence 325 AA;

Query Match 7.6%; Score 85.5; DB 1; Length 325;  
Best Local Similarity 24.9%; Pred. No. 0.13;  
Matches 42; Conservative 32; Mismatches 58; Indels 37; Gaps 9;

Qy 64 PLPLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPYGVFAGRDASRG 110  
Db 140 PLPCADRIDYTRDMVYQGNLEDIQKLNHLVINNKI--VKNIESAEWFTKLYKE 197  
Qy 111 -----GPYGVFAGRDASRLATFCLDKKAE-----LKDEYDLSLTAAQOETLSDWES 158  
Db 198 VTDFLPLUNVFSNRHLTK-IULTALEKEIIHLDDFLTDFTVMKLTSSRDKQIINLIS 256  
Qy 159 QFTFYHHVKKLLKEGEEPTVYSDEEPPKDES-SRKNVKAFFSGIS 203  
Db 257 KFDLNNHI-----IESKEDYEYSKKIKMRIDPLVNNQKVECAS-SIS 300

## RESULT 11

P60569  
ID P60569 standard; Protein; 1073 AA.  
AC P60569;  
DT 24-AUG-1991 (first entry)  
DE Sequence of the Ring-infected Erythrocyte Surface Antigen (RESA).  
KW Malaria vaccine; antigen; epitope.  
OS Plasmodium falciparum.  
PN WO8601802-A.  
PD 27-MAR-1986.  
PF 11-SEP-1985; 006960.  
PR 11-SEP-1984; AU-007067.  
PR 11-SEP-1984; AU-007086.  
PR 10-SEP-1985; AU-047326.  
PA (HALL-) HALL INST MED RES.  
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;  
DR WPI: 86-094065/14.  
DR N-PSDB: N60472.  
PT DNA coding for Plasmodium falciparum antigens - expressing poly:peptide(s) having antigenicity of RESA or FIRA antigens of P falciparum  
PS Disclosure; Fig 1; 55pp; English.  
CC The inventors claim a novel DNA molecule which comprises a nucleotide sequence corresp. to all or a portion of the base sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have antigenicity suitable for providing protective immunity against Plasmodium falciparum malarial infections.  
CC Sequence 1073 AA;

Query Match 7.0%; Score 79; DB 1; Length 1073;

Best Local Similarity 26.0%; Pred. No. 4;

Matches 34; Conservative 19; Mismatches 36; Indels 42; Gaps 7;

Qy 65 LPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPYGVFAGRDASRG 124  
Db 376 LPRLKRRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPYGVFAGRDASRG 124  
Qy 125 LATFCLDKKAEAL---KDEYDLSLTAAQOETLSDWESQFTFYHHVKKLLKEG---EET 177  
Db 408 LYT---DEDLLFLEKQKYMMLD--TSEESVEEENEHTVDDEHVEHTADDEHVEEP 462  
Qy 178 TVYSDE--EET 186  
Db 463 TVADDEHVEEP 473

## RESULT 12

R80220  
ID R80220 standard; peptide; 334 AA.

AC R80220;  
DT 29-APR-1996 (first entry)  
DE GRB-7 adaptor protein.  
KW PTK; oncogene; identification; detection; breast cancer; receptor;  
KW complex; adaptor; HER-2; GRB.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 2  
FT /note= "unspecified amino acid"  
FT misc\_difference 4  
FT /note= "unspecified amino acid"  
FT misc\_difference 5  
FT /note= "unspecified amino acid"  
PN WO9524205-A1.  
PD 14-SEP-1995.  
PF 07-MAR-1995; U02787.  
PR 07-MAR-1994; US-207575.  
PA (UNY-) UNIV NEW YORK MEDICAL CENT.  
PI Margolis BL;  
DR WPI: 95-328097/42.  
PT Identification of cpds. for modulating an oncogenic disorder esp.

PT	breast cancer - by exposing potential agents to a receptor protein					
PT	tyrosine kinase polypeptide/adaptor polypeptide complex					
PS	Disclosure; Fig 8B; 112pp; English.					
CC	Conserved motifs of the protein tyrosine kinase (PTK) catalytic domain may be complexed with an adaptor polypeptide to give a					
CC	receptor protein tyrosine kinase/adaptor protein (RptKp/AP) complex.					
CC	The adaptor protein is a member of the SH2 and SH3 contg. family of					
CC	adaptor proteins and is pref. a GRB-7 adaptor protein. A preferred					
CC	compound of the invention is an HPR2/GRB-7 complex. The complexes					
CC	can be used to screen for candidate compounds for modulating					
CC	oncogenic disorders in partic. breast cancer.					
SQ	Sequence 334 AA;					
	Query Match	6.8%;	Score 77;	DB 1;	Length 334;	
	Best Local Similarity	25.9%;	Pred. No. 1.3;			
	Matches 22;	Conservative 15;	Mismatches 30;	Indels 18;	Gaps 3	
QY	85 QDPRLIMAI----	NGKVFDVTKGRKYGP	PGYPVAGRDASRG	LATFCLDKEAL	KDEYD 140	
	:      :	:       :	:	:	:	
Db	176 KDRHLQIVADVNESN	VYVTGKRKLYGMPTDFG-----	FCVKPNKLNRGHK	222		
QY	141 DLSDLTAAQQETLSDWESQF-TFKY	164				
	: : : : :   :					
Db	233 GLHIFCSDEQSRCTCWLAAFRLFKY	247				
RESULT 13						
R80167						
ID	R80167 standard; peptide; 334 AA.					
AC	R80167;					
DT	22-APR-1996 (first entry)					
DE	Mouse signal transduction protein GRB-7 residues 95-428.					
KW	Signal transduction protein; growth factor receptor bound; BLM domain;					
KW	pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;					
KW	abnormal cell development; cell movement; breast cancer; atherosclerosis.					
OS	Mus musculus.					
PN	WO9525166-A1.					
PD	21-SEP-1995.					
PF	13-MAR-1995; U03452.					
PR	14-MAR-1994; US-212234.					
PR	08-JUN-1994; US-255785.					
PA	(UJNY-) UNIV NEW YORK MEDICAL CENT.					
PI	Labudry JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;					
PT	WPI; 95-336971/43.					
DR	Treating diseases involving abnormal signal transduction e.g. cancer					
PT	and psoriasis - by modulating interaction between e.g. epidermal					
PT	growth factor receptor and its ligand, also diagnosis and screening					
PT	of modulators					
PS	Claim 15; Fig 3; 102pp; English.					
CC	The amino acid sequence of the signal transduction protein, growth factor					
CC	receptor bound (GRB)-7 protein. This sequence covers from amino acids					
CC	95-428 of the full length protein. The protein contains a central BLM					
CC	domain and within this domain a pleckstrin domain (R80161). The central					
CC	protein is flanked by a proline-rich and an SH2 domain indicating that the					
CC	SH2 domain is involved in signal transduction. The SH2 domain has been shown					
CC	to bind to the HER2 receptor protein. The protein can be used to screen					
CC	for cpts. which can promote or interrupt interaction of proteins involved					
CC	in signal transduction, esp. in neuronal diseases, diseases involved with					
CC	abnormal cell development and defective cell movement, breast cancer,					
CC	atherosclerosis, etc.					
SQ	Sequence 334 AA;					
	Query Match	6.8%;	Score 77;	DB 1;	Length 334;	
	Best Local Similarity	25.9%;	Pred. No. 1.3;			
	Matches 22;	Conservative 15;	Mismatches 30;	Indels 18;	Gaps 3	
QY	85 QDPRLIMAI----	NGKVFDVTKGRFYGP	PGYPVAGRDASRG	LATFCLDKEAL	KDEYD 140	
	:      :	:         :	:	:	:	
Db	176 KDRHLQIVADVADVNESN	VYVTGKRKLYGMPTDFG-----	FCVKPNKLNRGHK	222		
QY	141 DLSDLTAAQQETLSDWESQF-TFKY	164				

PT growth factor receptor and its ligand, also diagnosis and screening  
PT of modulators  
PS Disclosure; Fig 3; 102pp; English.  
CC The amino acid sequence of the signal transduction protein, growth factor  
CC receptor bound (GRB)-7 protein. This sequence covers from amino acids  
CC 2-535 of the full length protein. The protein contains a central BLM  
CC domain and within this domain a pleckstrin domain (R80161). The central  
CC domain is flanked by a proline-rich and an SH2 domain indicating that the  
CC protein is involved in signal transduction. The SH2 domain has been shown  
CC to bind to the HER2 receptor protein. The protein can be used to screen  
CC for cpds. which can promote or interrupt interaction of proteins involved  
CC in signal transduction, esp. in neuronal diseases, diseases involved with  
CC abnormal cell development and defective cell movement, breast cancer,  
CC atherosclerosis, etc.  
SQ Sequence 534 AA;

```
Query Match      6.8; Score 77; DB 1; Length 534;  
Best Local Similarity 25.9%; Pred. No. 2.5;  
Matches 22; Conservative 15; Mismatches 30; Indels 18; Gaps 3;  
  
QY      85 QDPRLMAI-----NGKVFDVTKGRKFGPEGPVGFAGRDASRGATFCLDKAEALKDEYD 140  
       :||| | : : ||:||| || : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db      269 KOPRHLGVADVNESVVVTQGRKLGYMPTDFG-----FCVKPNKLIRNGHK 315  
  
QY      141 DLSDLTAAQQETLSDWESQF-TFKY 164  
       _ : :: : : : | | | |  
Db      316 GLHIFCSEDEQSRTCWLAAAFRLFXY 340
```

Search completed: April 19, 2000, 02:35:46  
Job time: 2607 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2000, 02:19:10 ; Search time 13.24 seconds  
(without alignments)  
220.956 Million cell updates/sec

Title: US-09-203-548-1

Perfect score: 1129

Sequence: 1 MAAEDVVATGADPSDLESGG.....SISXXYFAKSFVTVHXVFKT 220

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 133990 seqs, 13297546 residues

Total number of hits satisfying chosen parameters: 133990

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
  - 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*
  - 4: /cgn2\_6/ptodata/1/1aa/PCRU9\_COMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	976	86.4	195	2	US-08-960-022-6
2	81	7.2	535	1	US-08-167-035-10
3	81	7.2	535	1	US-08-208-887A-10
4	81	7.2	535	2	US-08-539-005-10
5	77	6.8	334	3	US-08-472-593-9
6	77	6.8	535	1	US-07-906-349A-10
7	76	6.7	502	2	US-08-960-022-10
8	75.5	6.7	1184	5	5254799-6
9	75.5	6.7	1188	5	5254799-7
10	74	6.6	251	1	US-08-152-019A-37
11	72.5	6.4	252	2	US-08-460-309-12
12	72.5	6.4	252	2	US-08-125-077-12
13	72.5	6.4	548	2	US-08-890-094-18
14	72	6.4	638	2	US-08-557-122A-38
15	71.5	6.3	3052	2	US-08-557-122A-26
16	71	6.3	709	1	US-07-814-964-7
17	71	6.3	709	1	US-08-258-442-7
18	71	6.3	709	1	US-08-328-809-2
19	71	6.3	709	4	PCT-US92-11107-7
20	70.5	6.2	252	1	US-08-152-019A-38
21	69.5	6.2	446	2	US-08-833-610-2
22	68	6.0	332	2	US-08-958-642-2
23	68	6.0	723	1	US-07-814-964-11
24	68	6.0	723	1	US-08-258-442-11
25	68	6.0	723	1	US-08-328-809-6
26	68	6.0	723	4	PCT-US92-11107-11
27	67.5	6.0	1019	1	US-08-271-364A-7
28	67.5	6.0	1019	2	US-08-222-715B-26
29	67	5.9	545	2	US-08-990-114-1

30	67	5.9	759	1	US-08-676-967-1	Sequence 1, Appli
31	67	5.9	759	1	US-08-676-974-1	Sequence 1, Appli
32	67	5.9	759	2	US-09-098-487-1	Sequence 1, Appli
33	66	5.8	452	1	US-08-317-880-2	Sequence 2, Appli
34	66	5.8	452	2	US-08-782-396-2	Sequence 2, Appli
35	66	5.8	811	1	US-08-136-743B-4	Sequence 4, Appli
36	66	5.8	1093	4	PCT-US94-04496-55	Sequence 55, Appli
37	65.5	5.8	410	1	US-07-955-905A-26	Sequence 26, Appli
38	65.5	5.8	536	2	US-08-890-094-2	Sequence 2, Appli
39	65.5	5.8	947	2	US-08-500-857A-4	Sequence 4, Appli
40	64.5	5.7	325	1	US-08-292-549-2	Sequence 2, Appli
41	64.5	5.7	325	4	PCT-US91-02207-2	Sequence 2, Appli
42	64.5	5.7	359	2	US-08-713-636-2	Sequence 2, Appli
43	64.5	5.7	605	2	US-08-687-956A-1	Sequence 1, Appli
44	64.5	5.7	812	1	US-08-446-794A-4	Sequence 4, Appli
45	64	5.7	366	1	US-08-700-359-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1  
US-08-960-022-6  
; Sequence 6, Application US/08960022  
; Patent No. 5976837  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/960,022  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 195 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-960-022-6

Query Match 86.4%; Score 976; DB 2; Length 195;  
Best Local Similarity 96.4%; Pred. No. 1.1e-103;  
Matches 187; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAAEDVATGADPSDLESGLLHEIFTSPLNLLLGICIFLLYKIVRGDOPAAASGDRTTT 60  
 Db 1 MAAEDVATGADPSDLESGLLHEIFTSPLNLLLGICIFLLYKIVRGDOPAAASGDDDD 60  
 QY 61 XPPPLPKRRDFTPAELRRRGGVODPRILMAINGKVFDTKGRKFGYGGPGYGVFAGRD 120  
 Db 61 EPPPLPKRRDFTPAELRRRGGVODPRILMAINGKVFDTKGRKFGYGGPGYGVFAGRD 120  
 QY 121 ASRGLATCLDKALKDEYDDLSLTAQAQETLSWESQFTFKYHHVGLKKEGEPTVY 180  
 Db 121 ASRGLATCLDKALKDEYDDLSLTAQAQETLSWESQFTFKYHHVGLKKEGEPTVY 180  
 QY 181 SDEEPKDESRKN 194  
 Db 181 SDEEPKDESARKN 194

## RESULT 2

US-08-167-035-10  
 ; Sequence 10, Application US/08167035  
 ; Patent No. 5618691  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schllessinger, Joseph  
 ; APPLICANT: Skolnick, Edward Y.  
 ; APPLICANT: Margolis, Benjamin L.  
 ; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
 ; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
 ; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: 10036-2711  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/167.035  
 ; FILING DATE: 16-DEC-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7683-062  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 535 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-167-035-10

Query Match 7.2%; Score 81; DB 1; Length 535;  
 Best Local Similarity 27.1%; Pred. No. 0.51;  
 Matches 23; Conservative 14; Mismatches 30; Indels 18; Gaps 3;  
 QY 85 QDPRILMAI----NGKVFDTKGRKFGYGGPGYGVFAGRDASRGATCLDKALKDEYD 140  
 Db 270 KDPRLQYADVNESNVIVTQGRKLYGMPDIFG-----FCVXPKNLRLNGHK 316  
 QY 141 DLSDLTAQAQETLSWESQF-TFKY 164  
 Db 317 GLHIFCSEDEQTRICWLAARLFKY 341

## RESULT 3

US-08-208-887A-10  
 ; Sequence 10, Application US/08208887A  
 ; Patent No. 5677421  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schllessinger, Joseph  
 ; APPLICANT: Skolnick, Edward Y.  
 ; APPLICANT: Margolis, Benjamin L.  
 ; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
 ; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
 ; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: 10036-2711  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/208.887A  
 ; FILING DATE: 11-MAR-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7683-063  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 535 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-208-887A-10

Query Match 7.2%; Score 81; DB 1; Length 535;  
 Best Local Similarity 27.1%; Pred. No. 0.51;  
 Matches 23; Conservative 14; Mismatches 30; Indels 18; Gaps 3;  
 QY 85 QDPRILMAI----NGKVFDTKGRKFGYGGPGYGVFAGRDASRGATCLDKALKDEYD 140  
 Db 270 KDPRLQYADVNESNVIVTQGRKLYGMPDIFG-----FCVXPKNLRLNGHK 316  
 QY 141 DLSDLTAQAQETLSWESQF-TFKY 164  
 Db 317 GLHIFCSEDEQTRICWLAARLFKY 341

## RESULT 4

US-08-539-005-10  
 ; Sequence 10, Application US/08539005  
 ; Patent No. 5858686  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schllessinger, Joseph  
 ; APPLICANT: Skolnick, Edward Y.  
 ; APPLICANT: Margolis, Benjamin L.  
 ; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
 ; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
 ; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:







```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-152-019A-37

Query Match      6.6%; Score 74; DB 1; Length 251;
Best Local Similarity 29.5%; Pred. No. 1.1;
Matches 44; Conservative 15; Mismatches 40; Indels

Qy   83 GVQDPRILMAIN-----GKVDVTKGR-KY-----CPEGDY
    ||| :|| || ||:| ||: | ||: | ||: | ||: | ||: |
Db   92 GVOYPN---SINLTGLGKAFDITYRLKFHTSRPESFAIYKTRTREDGFWIPY
    ||| :|| || ||:| ||: | ||: | ||: | ||: | ||: |
Qy   119 ---RDASRG-LATEFCLDEAL-KDEVDDLSDLTAAG--ETLSDWESQFTFK
    ||| :|| || ||:| ||: | ||: | ||: | ||: | ||: |
Db   149 ENTYSKANRGFTRTGDGEQAALCTDFSDLSPLTGNNVAFSTLEGRPSAYNEID
    ||| :|| || ||:| ||: | ||: | ||: | ||: | ||: |
Qy   171 LKEGEPT-----VYSDE--EEPK 187
    |:| | | | | | | | | | | | | | | | | | | | | |
Db   206 LQEWYTATDIRVTNLRLNTFGDEVENEKP 234
    ||| :|| || ||:| ||: | ||: | ||: | ||: | ||: |

RESULT 11
US-08-460-309-12
; Sequence 12, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Meros
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,309
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```



```
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/557,122A  
: FILING DATE: 11-DEC-1995  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Lambiris, Elias J.  
: REGISTRATION NUMBER: 33,728  
: REFERENCE/DOCKET NUMBER: 3980.204-US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 212-867-0123  
: TELEFAX: 212-878-9655  
: INFORMATION FOR SEQ ID NO: 26:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3052 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: US-08-557-122A-26
```

Query Match 6.3%; Score 71.5; DB 2; Length 3052;  
Best Local Similarity 30.0%; Pred. No. 73;  
Matches 24; Conservative 13; Mismatches 18; Indels 25; Gaps

```
Qy   131 DKALKDEY-----DDLSDLTAAQQFTLSDWESGFTPKYHHVKL-----L 171  
      ||::|||          |||::|::|::|::|::|::|::|::|::|::|  
Db   2868 DQAKVDELVDGGADDDLDLEAEAPDLEDDEEDDQKAVK----DELDQGAGDDDDL 2922  
  
Qy   172 KEGEPTVSDEEP--KDE 189  
      :| ||| : |::: |||  
Db   2924 EEAEPPMEEDDQKAVKE 2943  
      :| ||| : |::: |||
```

Search completed: April 19, 2000, 02:37:07  
Job time: 1077 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2000, 02:20:18 ; Search time 12.57 seconds  
(without alignments)  
825.557 Million cell updates/sec

Title: US-09-203-548-1  
Perfect score: 1129  
Sequence: 1 MAEDVATGADPSDLESG.....SISXXYFAKSVTVHXVKT 220

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR-62:\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	943.5	83.6	194	2 JC5260	progesterone membr
2	262.5	23.3	152	2 S65181	hypothetical prote
3	151	13.4	797	2 H71412	hypothetical prote
4	92	8.1	318	2 S71197	nitrate reductase
5	90.5	8.0	877	2 S72541	nitrate reductase
6	85.5	7.6	325	2 T00032	oxetanocin A resis
7	84.5	7.5	382	2 T08086	glutamate--ammonia
8	82.5	7.3	697	2 T16908	hypothetical prote
9	81	7.2	440	2 S60423	hypothetical prote
10	79.5	7.0	2186	2 T13169	tigargin - fruit fl
11	79	7.0	420	2 T01077	hypothetical prote
12	79	7.0	705	2 T12952	probable helicase
13	79	7.0	749	2 S00095	splicing regulator
14	79	7.0	749	2 A45294	Balbani ring 2.1
15	79	7.0	1073	1 A25526	ring-infected eryt
16	77.5	6.9	886	1 JC5085	replication licens
17	77	6.8	532	2 JC5412	epidermal growth f
18	77	6.8	535	2 C46243	epidermal growth f
19	76	6.7	334	2 S35791	DNA ligase (ATP) (
20	76	6.7	396	2 A27891	RepA protein - Bac
21	76	6.7	643	1 S32476	protein disulfide-
22	75.5	6.7	491	2 S31278	cytochrome P450 2B
23	74.5	6.6	570	2 T08778	hypothetical prote
24	74	6.6	221	2 T04238	hypothetical prote
25	74	6.6	645	1 A37323	protein disulfide-
26	74	6.6	1607	1 MMSB2	laminin gamma-1 ch
27	73.5	6.5	479	1 A32290	protein-tyrosine-p
28	73.5	6.5	491	2 A27117	cytochrome P450 2B
29	73.5	6.5	519	2 S66673	disulfide isomeras
30	73.5	6.5	685	2 S46309	initiator-binding

31 73.5 6.5 799 2 JH0797  
32 73.5 6.5 2796 2 JC4743  
33 73 6.5 598 2 E72733  
34 73 6.5 758 2 T10614  
35 72.5 6.4 181 2 B69051  
36 72.5 6.4 312 2 JC5962  
37 72.5 6.4 319 2 S44642  
38 72.5 6.4 548 2 I39175  
39 72.5 6.4 628 2 T02420  
40 72.5 6.4 770 1 P31VBC  
41 72.5 6.4 847 2 A49412  
42 72.5 6.4 873 2 T09582  
43 72.5 6.4 971 2 E69165  
44 72.5 6.4 1156 2 B70356  
45 72.5 6.4 1348 2 S27812

castor protein - f  
fatty-acid synthas  
probable membrane-  
hypothetical prote  
replication factor  
paired-box contain  
hypothetical prote  
SH2-domain protein  
hypothetical prote  
RNA-directed RNA p  
ribonucleoside-dip  
translation initia  
hypothetical prote  
chromosome assembl  
probable epidermal

ALIGNMENTS

RESULT 1

JC5260  
progesterone membrane binding protein - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 25-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 17-Mar-1999  
C:Accession: JC5260  
R:Falkenstein, E.; Meyer, C.; Eisen, C.; Scriba, P.C.; Wehling, M.  
Biochem. Biophys. Res. Commun. 229, 86-89, 1996  
A:Title: Full-length cDNA sequence of a progesterone membrane-binding protein from por  
A:Reference number: JC5260: MUID:97112407  
A:Contents: vascular smooth cell  
A:Accession: JC5260  
A:Molecule type: mRNA  
A:Residues: 1-194 <FAL>  
A:Cross-references: EMBL:X97114; NID:g1657408; PID:e257707; PID:g1657409  
F:14-42/Domain: transmembrane #status predicted <TMM>

Query Match 83.6%; Score 943.5; DB 2; Length 194;  
Best Local Similarity 93.3%; Pred. No. 6.8e-81;  
Matches 181; Conservative 4; Mismatches 8; Indels 1; Gaps 1;  
QY 1 MAEDVATGADPSDLESGGLLHEIFTSPLNLLLLGICIFLLYKIVRGDOPAAAGDRTTT 60  
DB 1 MAEDVATGADPSDLESGGLLHEIFTSPLNLLLLGICIFLLYKIVRGDOPAAAS-DSDDD 59  
QY 61 XPPPLRLKRRDFTPAELRRFDGVQDPRILMAINGVEDVTGKRKFYGPFGYVAGRD 120  
DB 60 EPPPLRLKRRDFTPAELRRFDGVQDPRILMAINGVEDVTGKRKFYGPFGYVAGRD 119  
QY 121 ASRGLATFCLDKREALKDEYDDLSLTAAQETLSDWESQFTFKYHHVHGKLLKEGEPTVY 180  
DB 120 ASRGLATFCLDKREALKDEYDDLSLTAAQETLNDWDSQFTFKYHHVHGKLLKEGEPTVY 179  
QY 181 SDEEKPDESARKN 194  
DB 180 SDEEKPDESARKN 193

RESULT 2

S65181  
hypothetical protein YPL170w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein P2515  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 05-Dec-1997  
C:Accession: S65181; S69429  
R:Purnelle, B.; Coster, F.; Goffeau, A.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S65154  
A:Accession: S65181  
A:Molecule type: DNA  
A:Residues: 1-152 <PUR>  
A:Cross-references: EMBL:273526; NID:g1370358; PID:e247050; PID:g1370359; MIPS:YPL170

```

QY      110  EGYGVF--AGRDASGLATFCLDKKALKEVDVDDLTLAAQOETLSDWESQFTFYHHV 167
      _  :  ||  ||:  ||::  :  |  |  :  ::
Db      539  GGADSI LLVAGTDAT-----DEFNAIHS LRKKQ-----LLEYVI 573
QY      168  GKLLKEGEE 176

```





A:Accession: C31926  
A:Molecule type: DNA  
A:Residues: 231-440 <FUE>  
A:Cross-references: GB:M22580; NID:g170977; PID:g806319  
R:James, C.M.; Indge, K.J.; Oliver, S.G.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64165  
A:Accession: S64180  
A:Molecule type: DNA  
A:Residues: 1-440 <JAW>  
A:Cross-references: EMBL:Z72686; NID:g1322761; PID:e243778; PID:g1322762; MIPS:YGL164c  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 7L

Query Match 7.28; Score 81; DB 2; Length 440;  
Best Local Similarity 21.28; Pred. No. 6;  
Matches 42; Conservative 15; Mismatches 69; Indels 72; Gaps 6;  
QY 68 LKRRDFTPAELRRFD-----GVQDPRILMAINGKVDYTKGRKFGV---PEGPYGV 115  
DB 166 LKASDFVSENNRKYKAWKSEKGTGDKELIKVQVPTFFVTLFSIFYNLISENNQSGI 225  
QY 116 FAGRDASRGLA-TFCLDKALKDEY-----DDLSDLTAAQOETLSDWESQTFKXHHV-- 168  
DB 226 VKMEDMKRARIYRLNSTGRKYDYFMKIEQDFNDGRYHEDDDKEDTPELAIDLNLHIKK 285  
QY 168 -----GKLLK----- 175  
DB 286 LFFSVSGKLLERQDSPVLVKIDRSDDKENESSEGKGLLDDITVAVSGYEAEDE 345  
QY 176 EPTVYSDEEPEKDESSRK 193  
DB 346 EDEDEDEEGKGDEERK 363

## RESULT 10

T13169  
tigrin - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
C:Accession: T13169  
R:Fogarty, F.J.; Fessler, L.I.; Bunch, T.A.; Yaron, Y.; Parker, C.G.; Nelson, R.E.; Brown  
Development 120, 1747-1758, 1994  
A:title: tigrin, a novel Drosophila extracellular matrix protein that functions as a ligand  
A:Reference number: 217625; MUID:95009506  
A:Accession: T13169  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2186 <FOG>  
A:Cross-references: EMBL:U09506; NID:g493069; PID:g493070; PIDN:AAA56998.1  
C:Function:  
A:Description: functions as a ligand for Drosophila alpha PS2 beta PS integrins

Query Match 7.08; Score 79.5; DB 2; Length 2186;  
Best Local Similarity 21.98; Pred. No. 65;  
Matches 40; Conservative 15; Mismatches 65; Indels 63; Gaps 6;  
QY 46 VRGDPAASGDRTTTTP-----PPLP-----RLKRRDFTPAELRRFDGV 84  
DB 2002 VEGDEPLEGTARPPNPAPVSTPAPLPYSRGGSGGFYRRQDYT----- 2052  
QY 85 QDPRILMAINGKVDYTKGRKFGPE-GPYGVFAGRDASRGLATFCLDKALKDEYD--- 141  
DB 2052 -----FNVPVGSASASGSGPTSSASASASLGNWRNPNASGDEPQQEVDLQ 2098  
QY 141 -----DLSDLTAAQOETLSDWESQTFKXHHVHVKLLKEGEPTVYSDEEPEKDE 189  
DB 2099 QQQIEELGWNEKLEDLGQQTQVEDTDWQ-----AEDLGQQQVQVQVEDDLHFDQFGHSS 2154  
QY 190 SSR 192

DB 2155 SSR 2157  
|||

## RESULT 11

T01077  
hypothetical protein T10P11.1 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01077  
R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffman, J.; Till, S.; de la Bastide  
hi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.  
submitted to the EMBL Data Library, November 1998  
A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.  
A:Reference number: Z14248  
A:Accession: T01077  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-420 <KAP>  
A:Cross-references: EMBL:AC002330; NID:g2262135; PID:g2262136  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Note: T10P11.1

Query Match 7.08; Score 79; DB 2; Length 420;  
Best Local Similarity 23.08; Pred. No. 8.7; Indels 54; Gaps 7;  
Matches 48; Conservative 21; Mismatches 86;  
QY 9 TGADPSDLESGLLHEIFTSPNLILLGLCIFLAYKIVRGDQPAASGDRTTTTPPPPLRL 68  
DB 25 SGSD-SDEELKGLSHEEYRQRLKMRKSAKCFWE-----NTPSP----- 66  
QY 69 KRRDFTPAELRRFDGVQDPRILMAINGKVDYTKGRKFGPYGPGYVAGRDASRGLATF 128  
DB 66 --RQDNEDSDENADEIQDN-----GGERDONSNGKGRKKSDESESDGLSRKRKSKS 118  
QY 129 CLDKREALKDEYDDLSDLTAAQOETLSDWESQ-----FTFKYHHVHVKL 170  
DB 119 SRSKRRRRKRSYD--SDSESGESDSEEDRRRRKRSKRSRFRKKRSHRKK- 176  
QY 171 LKEGEPTVYSDEEPEKDESSRKNNKAFS 199  
DB 176 -----TKYSDSESDSDSKAEISASS 197

## RESULT 12

T12952  
probable helicase T6H20.10 - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 20-Sep-1999  
C:Accession: T12952  
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;  
submitted to the Protein Sequence Database, July 1999  
A:Reference number: Z17586  
A:Accession: T12952  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-705 <CHO>  
A:Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.10  
A:Experimental source: cultivar Columbia; BAC clone T6H20  
C:Genetics:  
A:Gene: ATSP:T6H20.10  
A:Map position: 3  
A:Note: intron positions not resolved

Query Match 7.08; Score 79; DB 2; Length 705;  
Best Local Similarity 34.08; Pred. No. 17;  
Matches 34; Conservative 6; Mismatches 38; Indels 22; Gaps 5;

QY 91 MAINGK-----VFDVTGKRKFGPE-----GPYGVFAGRDASRGLATFCLDKA-----L 135

```

Query Match      7.0%; Score 79; DB 2; Length 749;
Best Local Similarity 24.1%; Pred. No. 18;
Matches 33; Conservative 19; Mismatches 49; Indels 36; Gaps

QY 62 PPLPRLKRDRFTPAELRRRDEGVDQPRILMAINGKRVDFDTKRGKFGYCPGCVFAGRDA 121
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 54 PPKKKKTAAMETTPAPPERPERPE-----KPEDLT-----NPGGP--VDGGNG 98

```

```

Qy 122 SRGLATFCLDKALKDEYDLDLSLTAAQOETLSDWSQSOTFFKYHHVGVKLLKEGEPT --- 179
Db 99 -----DRYDEYDDDYDEYYDDEYDEYEYEDESTTAGVKP TGKK 147
Qy 179 -----VYSDEEPEKD 188
Db 148 GKOKKNKVDSDEETE 164

```

C;Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology  
C;Keywords: surface antigen; tandem repeat  
F;1-65/Domain: signal sequence #status predicted <SIG>  
F;66-1073/Product: ring-infected erythrocyte surface antigen #status predicted <MAT>  
F;523-587/Domain: dnaJ amino-terminal homology <DNJ>  
F;891-1073/Region: 4-residue repeats

Db 463 TVADDEHVEEP 473

Search completed: April 19, 2000, 02:37:31  
Job time: 1033 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2000, 02:35:47 ; Search time 10.22 Seconds  
(without alignments)

642.884 Million cell updates/sec

Title: US-09-203-548-1

Perfect score: 1129

Sequence: 1 MAEDVVATGADPSDLEGG.....SISXXYFAKSFVTVHXVFKT 220

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	978	86.6	194	1 MAPR_HUMAN	O00264 homo sapien
2	945	83.7	194	1 MAPR_MOUSE	O55022 sus musculus
3	938.5	83.1	193	1 MAPR_PIG	Q95250 sus scrofa
4	927	82.1	194	1 MAPR_RAT	P70580 rattus norv
5	501	44.4	223	1 SRP6_HUMAN	O15173 homo sapien
6	277	24.5	166	1 YRBF_SCHPO	Q13995 schizosacch
7	262.5	23.3	152	1 YP70_YEAST	Q12091 saccharomyc
8	92	8.1	318	1 NIA_CHLVU	Q01170 chlorella v
9	84.5	7.5	382	1 GLN1_CHLRE	Q42688 chlamydomon
10	82.5	7.3	697	1 SSRP_CAEEL	P41848 caenorhabdi
11	82	7.3	760	1 RESA_PLAFN	P13831 plasmodium
12	81	7.2	440	1 YGR4_YEAST	P53107 saccharomyc
13	79	7.0	749	1 SWAP_CAEEL	Q10580 caenorhabdi
14	79	7.0	1073	1 RESA_PLAFF	P13830 plasmodium
15	77.5	6.9	886	1 MCW2_XENLA	P55861 xenopus lae
16	77	6.8	532	1 GRB7_HUMAN	Q14451 homo sapien
17	77	6.8	535	1 GRB7_MOUSE	Q03160 mus musculus
18	76.5	6.8	335	1 MALR_CLOBU	O05103 clostridium
19	76	6.7	396	1 REPA_BACSU	P13962 bacillus su
20	76	6.7	643	1 ER72_RAT	P38659 rattus norv
21	75.5	6.7	491	1 CPB5_RABIT	P12789 oryctolagus
22	75.5	6.7	895	1 MCW2_HUMAN	P49736 homo sapien
23	74	6.6	645	1 ER72_HUMAN	P13667 homo sapien
24	74	6.6	863	1 MYSP_TAESO	P35418 taenia soli
25	74	6.6	1607	1 LMGI_MOUSE	P02468 mus musculus
26	74	6.6	1722	1 YP93_CAEEL	O09475 caenorhabdi
27	73.5	6.5	479	1 MP1P_DROME	P20483 drosophila
28	73.5	6.5	519	1 PD1R_HUMAN	Q14554 homo sapien
29	72.5	6.4	319	1 YPT5_CAEEL	P41883 caenorhabdi
30	72.5	6.4	770	1 RRP3_INBAC	P13875 influenza b
31	72.5	6.4	806	1 R1R1_PLAF4	P50648 plasmodium
32	72.5	6.4	814	1 IF3X_HUMAN	P55884 homo sapien
33	72.5	6.4	1959	1 MYN1_CHICK	P14105 gallus gall
34	72.5	6.4	2483	1 PCX_DROME	P18490 drosophila

35 72 6.4 638 1 ER72\_MOUSE  
36 71.5 6.3 327 1 RRPB\_PIRYV  
37 71.5 6.3 532 1 YHB7\_YEAST  
38 71.5 6.3 638 1 CIKE\_RAT  
39 71 6.3 429 1 UL88\_HCMVA  
40 71 6.3 490 1 YPLC\_CLOPE  
41 71 6.3 709 1 SSRP\_HUMAN  
42 70.5 6.2 671 1 ANX6\_CHICK  
43 70.5 6.2 1609 1 LMGI\_HUMAN  
44 70 6.2 1976 1 MYSQ\_HUMAN  
45 69.5 6.2 204 1 RUBY\_METJA

#### ALIGNMENTS

RESULT 1  
MAPR\_HUMAN  
ID MAPR\_HUMAN STANDARD; PRT; 194 AA.  
AC O00264;  
DT 15-DEC-1999 (Rel. 39, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE MEMBRANE ASSOCIATED PROGESTERONE RECEPTOR COMPONENT.  
GN HPR6.6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 98368853.  
RA GERDES D., WEHLING M., LEUBE B., FALKENSTEIN E.;  
RT "Cloning and tissue expression of two putative steroid membrane  
RT receptors";  
RL Biol. Chem. 379:907-911(1998).  
CC -!- FUNCTION: RECEPTOR FOR PROGESTERONE (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: MICROSMAL; MEMBRANE-BOUND (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE MAPR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y12711; CAA73248.1; -  
KW Receptor; Steroid-binding; Transmembrane; Microsome.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT TRANSMEM 24 42 POTENTIAL.  
SQ SEQUENCE 194 AA; 21540 MW; A06BA160 CRC32;  
  
Query Match 86.6%; Score 978; DB 1; Length 194;  
Best Local Similarity 96.9%; Pred. No. 2.7e-85;  
Matches 187; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 AEDVVATGADPSDLESGGLLHEIFTSPLNLLGLICIFLLYKIVRGDQPAASGDRITTX 61  
Db 1 AEDVVATGADPSDLESGGLLHEIFTSPLNLLGLICIFLLYKIVRGDQPAASGDSDDDE 60  
QY 62 PPPLRLKRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPFGYGFAGRDA 121  
Db 61 PPPLRLKRDFTPAELRRFDGVQDPRILMAINGKVFDTKGRKFYGPFGYGFAGRDA 120  
QY 122 SRGLATFCLDKALKDEYDDLSLTAAQOETLSDWESQFTFKYHHVGLLKEGEPPIVYS 181  
Db 121 SRGLATFCLDKALKDEYDDLSLTAAQOETLSDWESQFTFKYHHVGLLKEGEPPIVYS 180  
QY 182 DEEPEKDESSRKN 194  
|||||:|||||:|||||

Db 181 DEEPKDESARKN 193

## RESULT 2

MAPR\_MOUSE  
ID MAPR\_MOUSE STANDARD; PRT; 194 AA.  
AC OS5022;  
DT 15-DEC-1999 (Rel. 39, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE MEMBRANE ASSOCIATED PROGESTERONE RECEPTOR COMPONENT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-TESTIS;  
RA KWON S., LUNN R.M., O'BRIEN D.A., BELL D.A., EDDY E.M.;  
RT "The expression of a putative membrane associated progesterone  
receptor component in the mouse testis and epididymis.";  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: RECEPTOR FOR PROGESTERONE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: MICROSOMAL; MEMBRANE-BOUND (BY  
SIMILARITY). BELONGS TO THE MAPR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF042491; AAB97466.1;  
KW Receptor; Steroid-binding; Transmembrane; Microsome.  
FT INIT\_MET 0 BY SIMILARITY.  
FT TRANSMEM 24 42 POTENTIAL.  
SQ SEQUENCE 194 AA; 21553 MW; B5047CC6 CRC32;

Query Match 83.7%; Score 945; DB 1; Length 194;

Best Local Similarity 92.7%; Pred. No. 3.5e-82;

Matches 179; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 AADVATGADPSDLESGLLHEIFTSPLNLLGLCFLYLYKIVRGDQPAASGDRRTTX 61  
Db 1 AADVATGADPSDLESGLLHEIFTSPLNLLGLCFLYLYKIVRGDQPAASGDRRTTX 60  
QY 62 PPPLPKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPYGVFAGRDA 121  
Db 61 PPPLPKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPYGVFAGRDA 120  
QY 122 SRGLATFCLDKALKDEYDLSLTAAOQETLSDWESQTFKYHHVGLKKEGEPTVYS 181  
Db 121 SRGLATFCLDKALKDEYDLSLTAAOQETLSDWESQTFKYHHVGLKKEGEPTVYS 180  
QY 182 DEEPKDESSRKN 194  
Db 181 DEEPKDETARKN 193

## RESULT 3

MAPR\_PIG  
ID MAPR\_PIG STANDARD; PRT; 193 AA.  
AC Q95250;  
DT 15-DEC-1999 (Rel. 39, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE MEMBRANE ASSOCIATED PROGESTERONE RECEPTOR COMPONENT.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE-VASCULAR SMOOTH MUSCLE;  
RX MEDLINE; 97112407  
RA FALKENSTEIN E., MEYER C., EISEN C., SCRIBA P.C., WEHLING M.;  
RT "Full-length cDNA sequence of a progesterone membrane-binding protein  
from porcine vascular smooth muscle cells.";  
RL Biochem. Biophys. Res. Commun. 229:86-89(1996).  
RN [2]  
RP SEQUENCE OF 1-20, AND CHARACTERIZATION.  
RC TISSUE=LIVER;  
RX MEDLINE; 96370823.  
RA MEYER C., SCHMID R., SCRIBA P.C., WEHLING M.;  
RT "Purification and partial sequencing of high-affinity progesterone-  
binding site(s) from porcine liver membranes.";  
RL Eur. J. Biochem. 239:726-731(1996).  
CC -1- FUNCTION: RECEPTOR FOR PROGESTERONE.  
CC -1- SUBUNIT: HETERODIMER OR HETEROTRIMER (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: MICROSOMAL; MEMBRANE-BOUND.  
CC -1- SIMILARITY: BELONGS TO THE MAPR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X99714; CAA68050.1;  
KW Receptor; Steroid-binding; Transmembrane; Microsome.  
FT INIT\_MET 0  
FT TRANSMEM 24 42 POTENTIAL.  
SQ SEQUENCE 193 AA; 21478 MW; 1FD46521 CRC32;

Query Match 83.1%; Score 938.5; DB 1; Length 193;

Best Local Similarity 93.3%; Pred. No. 1.4e-81;

Matches 180; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 2 AADVATGADPSDLESGLLHEIFTSPLNLLGLCFLYLYKIVRGDQPAASGDRRTTX 61  
Db 1 AADVATGADPSDLESGLLHEIFTSPLNLLGLCFLYLYKIVRGDQPAAS-DSDDDE 59  
QY 62 PPPLPKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPYGVFAGRDA 121  
Db 60 PPPLPKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPYGVFAGRDA 119  
QY 122 SRGLATFCLDKALKDEYDLSLTAAOQETLSDWESQTFKYHHVGLKKEGEPTVYS 181  
Db 120 SRGLATFCLDKALKDEYDLSLTAAOQETLSDWESQTFKYHHVGLKKEGEPTVYS 179  
QY 182 DEEPKDESSRKN 194  
Db 180 DEEPKDESARKN 192

## RESULT 4

MAPR\_RAT  
ID MAPR\_RAT STANDARD; PRT; 194 AA.  
AC P70580; 070606;  
DT 15-DEC-1999 (Rel. 39, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE MEMBRANE ASSOCIATED PROGESTERONE RECEPTOR COMPONENT (ACIDIC 25 KD  
PROTEIN) (25-DX).  
CN LEWIS OR 25DX.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;  
RX MEDLINE; 97158736.

RA SELWIN O., LUCIER G.W., CLARK G.C., TRITSCHER A.M.,  
 RA VANDEN HEUVEL J.P., GASTEL J.A., WALKER N.J., SUTTER T.R., BELL D.A.;  
 RT "Isolation and characterization of a novel gene induced by 2,3,7,8-  
 RT tetrachlorodibenzo-p-dioxin in rat liver.";  
 RN Carcinogenesis 17:2609-2613(1996).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FISHER 344; TISSUE=LIVER;  
 RA NOELTE I., SOHN K., WEGEHLING S., WIELAND F.;  
 RT "Rat homolog to a putative progesterone binding protein: molecular  
 RT characterization and localization.";  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE OF 1-14.  
 RC STRAIN-WISTAR; TISSUE=LIVER;  
 RA HUBBARD M.J., MCHUGH N.J.;  
 RT "Acidic 25-kDa protein in rat liver microsomes.";  
 RL Submitted (JUL-1999) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: RECEPTOR FOR PROGESTERONE (BY SIMILARITY). MAY BE  
 CC IMPLICATED IN TCDD IMMUNOTOXICITY.  
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL; MEMBRANE-BOUND (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN LUNG, LIVER,  
 CC KIDNEY AND BRAIN, LOW IN TESTIS AND SPLEEN. NOT EXPRESSED IN HEART  
 CC AND SKELETAL MUSCLE.  
 CC -1- INDUCTION: BY DIOXIN.  
 CC -1- SIMILARITY: BELONGS TO THE MAPR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; U63315; AAB07125.1; .  
 DR EMBL; AJ005837; CAA06732.1; .  
 KW Receptor; Steroid-binding; Transmembrane; Microsome.  
 FT INIT\_MET 0  
 FT TRANSMEM 24 42  
 FT CONFLICT 160 166  
 FT CONFLICT 171 173  
 FT CONFLICT 177 177  
 FT CONFLICT 185 194  
 FT PKDEARKSD -> QMRLLRVTAVSGAYLFLFAKSFV  
 FT TFQSVFTTW (IN REF. 1).  
 SQ SEQUENCE 194 AA; 21467 MW; 86756116 CRC32;  
 -----  
 Query Match 82.1%; Score 927; DB 1; Length 194;  
 Best Local Similarity 90.7%; Pred. No. 1.7e-80;  
 Matches 175; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 AAEDVVATGADPSLEGGLLHEIFTSPLNLLGLCIFLLYKIVRGDPAASGDRITTX 61  
 Db 1 AAEDVVATGADPSLEGGLLHEIFTSPLNLLGLCIFLLYKIVRGDPAASGDRITTX 60  
 QY 62 PPLPLRRRDTPAELRRFDGVDQDPRILMAINGKVFDTKGRKFGPGPGYGVFAGRA 121  
 Db 61 PPLPLRRRDTPAELRRFDGVDQDPRILMAINGKVFDTKGRKFGPGPGYGVFAGRA 120  
 QY 122 SGLATFLDKALKDEYDDLSLTAQOETLSDWESQTFKYHHVGLLKEGEPTVYS 181  
 Db 121 SGLATFLDKALKDEYDDLSLTAQOETLSDWESQTFKYHHVGLLKEGEPTVYS 180  
 QY 182 DEEPKDESSRKN 194  
 Db 181 DEEPKDESSRKN 193  
 RESULT 5  
 SRD6\_HUMAN  
 ID SRD6\_HUMAN STANDARD; PRT; 223 AA.  
 AC O15173;

DT 15-DEC-1999 (Rel. 39, Created)  
 DT 15-DEC-1999 (Rel. 39, Last sequence update)  
 DT 15-DEC-1999 (Rel. 39, Last annotation update)  
 DE STEROID RECEPTOR PROTEIN DG6.  
 GN DG6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE; 98368853.  
 RA GERDES D., WEHLING M., LEUBE B., FALKENSTEIN E.;  
 RT "Cloning and tissue expression of two putative steroid membrane  
 RT receptors.";  
 RL Biol. Chem. 379:907-911(1998).  
 CC -1- FUNCTION: IS A RECEPTOR FOR STEROIDS (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE MAPR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AJ002030; CAA05152.1; .  
 KW Receptor; Steroid-binding; Transmembrane; Microsome.  
 FT TRANSMEM 42 66  
 FT SEQUENCE 223 AA; 23818 MW; 45CA4CDE CRC32;  
 SQ SEQUENCE 223 AA; 23818 MW; 45CA4CDE CRC32;  
 -----  
 Query Match 44.4%; Score 501; DB 1; Length 223;  
 Best Local Similarity 56.9%; Pred. No. 3.2e-40;  
 Matches 95; Conservative 30; Mismatches 38; Indels 4; Gaps 1;  
 QY 30 LNLGLLGLCIFLLYKI-----VRGDQPAASGDRITTPPLPLRRRDTTPAELRRFDG 85  
 Db 56 LNLVALVLLGALVLRVWRRGLGAGAGAGEESPATSLPRMKRDFSLQLRQYDGR 115  
 QY 86 DPRILMAINGKVFDTKGRKFGPGYGVFAGRASGLATFLDKALKDEYDDLSL 145  
 Db 116 NPRLILAVNGKVFDTKGRKFGPGYGVFAGRASGLATFLDKALKDEYDDLSL 175  
 QY 146 TAAQOETLSDWESQTFKYHHVGLLKEGEPTVYSDEEPKDESSR 192  
 Db 176 NAVQMESVREWEWQFKYDYVGRLLKPGEEPSEYDDEETKDHKK 222  
 -----  
 RESULT 6  
 YEBF\_SCHPO  
 ID YEBF\_SCHPO STANDARD; PRT; 166 AA.  
 AC O13995;  
 DT 15-DEC-1999 (Rel. 39, Created)  
 DT 15-DEC-1999 (Rel. 39, Last sequence update)  
 DT 15-DEC-1999 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 18.9 KD PROTEIN C26H5.15 IN CHROMOSOME I.  
 GN SPAC26H5.15.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MAPR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE STRUCTURE-SPECIFIC RECOGNITION PROTEIN 1 (SSRP1)
DE (RECOMBINATION SIGNAL SEQUENCE RECOGNITION PROTEIN).
GN T20B12.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC WATERSTON R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -!- SIMILARITY: BELONGS TO THE SSRP FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10401; AAA19061.1; -
CC HSSP; P07155; 1HME.
CC WORMPEP; T20B12.8; CE01414.
CC PFAM; PF00505; HMG_box; 1.
CC Hypothetical protein; Nuclear protein; DNA-binding.
CC DNA_BIND 556 HMG BOX.
CC SEQUENCE 697 AA; 78635 MW; E78A3DED CRC32;
CC -----
Query Match 7.3%; Score 82.5; DB 1; Length 697;
Best Local Similarity 21.1%; Pred. No. 4.2;
Matches 46; Conservative 28; Mismatches 91; Indels 53; Gaps 8;
QY 4 EDVVATGADPSDLESGLLHEITSPINLLLLGICILLYKIVRGDQPAASGDRTTTXXPP 63
DB 485 ESSEGTSDEPDDEYDSG-----SEQDSSGSGSEPDSEQ 518
QY 64 PLPLRLKRRDTPAELRFRDVGQDPRILMAINGKVFDTGKRKFGPGPYGVFAGRDASR 123
DB 519 DVPS-KRKKEGPKREKREKK-----EGK-----KGGKKDKDNPAP-----KRATSA 562
QY 124 GLATFCLDKALKDEYDDLSDLT---AAQDETJSD-----WESGFTFYHHVHGKLLKEGE 175
DB 563 YMQWFLASRNELKEDGSDVADVAKKGAKWKMTSSDDKKWEEKAEEDKSRYEKMEYR 622
QY 176 EPTVYSDEEPKDESSRKNVKAFGSGISXXYFAKSFVT 213
DB 623 KNGPPSSSKPP---SSSKTSKSGSPSSSKAISKEYIS 657
RESULT 11
RESA_PLAFN
ID RESA_PLAFN STANDARD; PRT; 760 AA.
AC PL3831;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-MAY-1992 (Rel. 22, Last annotation update)
DE RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN (FRAGMENT).
GN RESA.
OS Plasmodium falciparum (isolate NF7 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 85295381.
RA COWMAN A.F., COPPEL R.L., SAINT R.B., FAVALORO J., CREWTER P.E.,
RA STAHL H.-D., BLANCO A.E., BROWN G.V., ANDERS R.F., KEMP D.J.;
RT "The ring-infected erythrocyte surface antigen (RESA) polypeptide of
RT Plasmodium falciparum contains two separate blocks of tandem repeats

```





KW Transcription regulation: RNA-binding; mRNA splicing; Repeat.  
FT DOMAIN 8 124 DRY CEERYL.  
FT PROSITE: 2 X REPEATS OF THE SURP MOTIF.  
FT REPEAT 166 431 SURP MOTIF 1.  
FT REPEAT 166 209 SURP MOTIF 2.  
FT REPEAT 331 431 SURP MOTIF 2.  
FT DOMAIN 688 749 ARG/LYS/SER-RICH (HIGHLY BASIC).  
SQ SEQUENCE 749 AA; 86309 MW; 587FBC82 CRC32;

Query Match 7.0%; Score 79; DB 1; Length 749;  
Best Local Similarity 23.6%; Pred. No. 9.8;  
Matches 54; Conservative 25; Mismatches 74; Indels 76; Gaps 10;

QY 6 VVATGADPSDLGGGLLHEIFTSPLMLLLGICFLLYKIVRGDQPAASGDRITTPPLP 65  
DB 454 VLNSAPPASVSPG-----PSSLSMLNL-----STREPPL 484  
QY 66 PLKRRDFTPAELRRFD-----GVQDPRLMAINGKV-----FDVTKGRKFGPEGYPG 114  
DB 485 NRRQRRLLDSS--RLDESITPEGVDPITMLQIPKSVSTPANLDILK-----TPIS 534  
QY 115 VFAGRDASRLATFCLDKKALKDEYDLSLTA-----AQQETLSD 155  
DB 535 FSLRNDEPRDESFRDPDL--DETAGPSDTTANFSDISGLFPPTPPVIPPSTQMQVDR 592  
QY 156 WESQFTKYHHVKKLKEGEPTVYSDSEPK--DESRKNVKAFIGSIS 203  
DB 593 KEKARIF-----MEKLQEKKAKKLQEEERSKLEETRKAERKISLSLS 637

## RESULT 14

RESA\_PLAFF  
ID RESA\_PLAFF STANDARD; PRT; 1073 AA.  
AC P13030;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN PRECURSOR.  
GN RESA.  
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87066710.  
RA FAVALORO J.M., COPPEL R.L., CORCORAN L.M., FOOTE S.J., BROWN G.V.,  
RT ANDERS R.F., KEMP D.J.;  
RL "Structure of the RESA gene of Plasmodium falciparum.";  
RL Nucleic Acids Res. 14:8265-8277(1986).  
CC -!- FUNCTION: RESA MAY DISRUPT THE NORMAL INTERMOLECULAR INTERACTIONS  
CC OF THE CYTOPLASMIC DOMAIN OF BAND 3 AND THEREBY FACILITATE THE  
CC INVAGINATION OF THE RED CELL MEMBRANE WHICH IS NECESSARY FOR THE  
CC FORMATION OF THE PARASITOPHOUS VACUOLE.  
CC -!- SUBCELLULAR LOCATION: PROBABLY LOCATED ON THE CYTOPLASMIC FACE OF  
CC THE MEMBRANE WHERE IT ASSOCIATES WITH COMPONENTS OF THE MEMBRANE  
CC SKELETON.  
CC -!- PTM: THE TYR RESIDUES IN THE VARIANT TETRAMERIC SEQUENCES IN THE  
CC RESA REPEAT ARE POSSIBLY PHOSPHORYLATED (BY HOMOLOGY WITH BAND 3).  
CC -!- SIMILARITY: THE N-TERMINAL SEQUENCE OF BAND 3 SHOWS HOMOLOGY WITH  
CC THE REPEAT SEQUENCES OF RESA.  
CC -!- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X04572; CAA28241.1; -;  
DR EMBL; X05182; CAA28817.1; -;  
DR EMBL; A00661; CAA00077.1; -;  
PIR; A25526; A25526.

DR HSSP; P08622; 1XBL.  
DR PROSITE: PS00636; DNAJ\_1; 1.  
DR PROSITE: PS00076; DNAJ\_2; 1.  
DR PFAM; PF00226; DnaJ; 1.  
KW Malaria; Antigen; Glycoprotein; Repeat; Signal.  
FT SIGNAL 1 65  
FT CHAIN 66 1073 RING-INFECTED ERYTHROCYTE SURFACE  
FT ANTIGEN.  
FT DOMAIN 436 504  
FT DOMAIN 521 589  
FT DOMAIN 891 1073  
FT CARBOHYD 71 71  
FT CARBOHYD 639 639  
FT CARBOHYD 773 773  
FT CARBOHYD 777 777  
FT VARIANT 378 378  
SQ SEQUENCE 1073 AA; 124906 MW; E08699C5 CRC32;

Query Match 7.0%; Score 79; DB 1; Length 1073;

Best Local Similarity 26.0%; Pred. No. 16;  
Matches 34; Conservative 19; Mismatches 36; Indels 42; Gaps 7;

QY 65 LPLKRRDFTPAELRRFDGVQDPRLMAINGKVFDVTKGRKFGPEGYPGVFAGRDASRG 124  
DB 376 LPSL-RASITNSAINYYDVTVD-----GVYLDHETSDA 407  
QY 125 LATFCLDKKAL-----KDEYDLSLTAQQTSLDWSQFTFYHHVKKLKEG---EEP 177  
DB 408 LYT---DEDLLFLDKQKYMMLD--TSEESVEENEETVDDERHEEHTADDERVEEP 462  
QY 178 TVYSDE---EEP 186  
DB 463 TVADDERVEEP 473

## RESULT 15

MC2\_XENLA  
ID MC2\_XENLA STANDARD; PRT; 886 AA.  
AC P55861;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DNA REPLICATION LICENSING FACTOR MC2 (X.MCM2).  
GN MCM2.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae;  
OC Xenopus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-OOCYTE;  
RX MEDLINE; 97074651.  
RA MIYAKE S., SAITO I., KOBAYASHI H., YAMASHITA S.;  
RT "Identification of two Xenopus laevis genes, xMC2 and xCDC46, with  
RL sequence homology to MCM genes involved in DNA replication.";  
RL Gene 175:71-75(1996).  
CC -!- FUNCTION: ACT AS A FACTOR THAT LICENSE THE DNA FOR ONE AND ONLY  
CC ONE ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR THE ENTRY IN  
CC S PHASE AND FOR CELL DIVISION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -!- SIMILARITY: BELONGS TO THE MCM FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D63919; BAA09948.1; -;  
DR PROSITE; PS00847; MCM\_1; 1.

DR PROSITE; PSS0051; MCM\_2; 1.  
DR PFAM; PF00493; MCM; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW DNA replication; Zinc-finger; ATP-binding; Cell cycle.  
FT ZN\_FING 314 340 C4-TYPE (POTENTIAL).  
FT DOMAIN 458 665 MCM.  
FT NP\_BIND 508 515 ATP (POTENTIAL).  
SQ SEQUENCE 886 AA; 100376 MW; C79D0ED6 CRC32;

Query Match 6.98; Score 77.5; DB 1; Length 886;  
Best Local Similarity 32.8%; Pred. No. 17;  
Matches 22; Conservative 12; Mismatches 26; Indels 7; Gaps 3;  
QY 130 LDK---EALKDEYDDLTLTAQOETLSDWESQFTFKYHHVGLKLLKEGEEPTVYSDEEEP 186  
DB 72 LDRIYVEGLDDE-EDVEDLTASQREAA---EQSMRMRDREMGRELGRMRRGLLYDSDEE 127  
QY 187 KDESSRK 193  
DB 128 EDRPARK 134

Search completed: April 19, 2000, 03:19:47  
Job time: 2640 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 19, 2000, 02:32:32 ; Search time 13.52 seconds  
(without alignments)  
1128.218 Million cell updates/sec

Title: US-09-203-548-1  
Perfect score: 1129  
Sequence: 1 MAEDVAVGADPSDLESGS.....SISXXYFAKSVTVHXVKFT 220

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	302.5	26.8	964	5 Q10461	Q10461 caenorhabdi
2	244	21.6	253	10 Q9XPM5	Q9XPM5 arabidopsis
3	241	21.3	220	10 Q9XFM6	Q9XFM6 arabidopsis
4	151	13.4	797	10 Q23350	Q23350 arabidopsis
5	143.5	12.7	326	5 Q9XXA7	Q9XXA7 caenorhabdi
6	90.5	8.0	877	10 Q42497	Q42497 chlorella v
7	86.5	7.7	2380	5 Q22896	Q22896 caenorhabdi
8	85.5	7.6	325	2 Q24771	Q24771 bacillus me
9	82	7.3	124	3 Q94391	Q94391 schizosacch
10	81.5	7.2	692	11 Q61033	Q61033 mus musculu
11	81.5	7.2	1451	5 Q01737	Q01737 caenorhabdi
12	80.5	7.1	979	2 Q9XAS7	Q9XAS7 streptococc
13	79.5	7.0	2186	5 Q23984	Q23984 drosophila
14	79	7.0	420	10 Q22758	Q22758 arabidopsis
15	79	7.0	749	5 Q23804	Q23804 chironomus
16	78	6.9	224	11 Q61028	Q61028 mus musculu
17	78	6.9	342	11 Q61032	Q61032 mus musculu
18	78	6.9	379	11 Q61031	Q61031 mus musculu
19	78	6.9	411	11 Q61030	Q61030 mus musculu
20	78	6.9	451	11 Q61029	Q61029 mus musculu

21	77.5	6.9	149	10 Q9ZSP7	Q9ZSP7 petunia hyb
22	77.5	6.9	886	13 Q42588	Q42588 xenopus lae
23	77	6.8	447	4 Q9Y220	Q9Y220 homo sapien
24	76.5	6.8	653	5 Q22553	Q22553 caenorhabdi
25	76	6.7	334	5 Q23694	Q23694 crithidia f
26	75.5	6.7	734	5 Q94056	Q94056 caenorhabdi
27	75	6.6	461	5 Q44628	Q44628 caenorhabdi
28	74.5	6.6	432	5 P91584	P91584 ciona intes
29	74.5	6.6	562	4 Q75759	Q75759 homo sapien
30	74.5	6.6	577	4 Q9Y218	Q9Y218 homo sapien
31	74	6.6	803	5 Q17493	Q17493 caenorhabdi
32	74	6.6	2356	5 Q46008	Q46008 caenorhabdi
33	73.5	6.5	276	5 Q44512	Q44512 caenorhabdi
34	73.5	6.5	322	2 Q68594	Q68594 pseudomonas
35	73.5	6.5	448	4 Q15468	Q15468 homo sapien
36	73.5	6.5	491	6 Q29516	Q29516 oryctolagus
37	73.5	6.5	816	5 Q27537	Q27537 caenorhabdi
38	73.5	6.5	858	5 P90545	P90545 entamoeba h
39	73.5	6.5	2408	5 Q22184	Q22184 caenorhabdi
40	73.5	6.5	2796	2 Q48926	Q48926 mycobacteri
41	73	6.5	598	1 Q9YF35	Q9YF35 aeropyrum p
42	73	6.5	663	5 Q24967	Q24967 giardia lam
43	73	6.5	783	3 Q13956	Q13956 schizosacch
44	73	6.5	793	5 Q22047	Q22047 caenorhabdi
45	73	6.5	1045	3 Q74473	Q74473 schizosacch

## ALIGNMENTS

### RESULT 1

Q10461 ID Q10461 PRELIMINARY; PRT; 964 AA.  
AC Q10461;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE HYPOTHETICAL 106.7 KD PROTEIN IN K07E3.6 IN CHROMOSOME X.  
GN K07E3.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA LATREILLE P., GATTUNG S.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: WEAK, TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2 CC ATPASES).  
DR EMBL; U41552; AAC69103.1; -;  
DR WORMPEP; K07E3.6; CE04722.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
DR PFAM; PF00122; E1-E2\_ATPase; 4.  
KW Hypothetical protein.  
SQ SEQUENCE 964 AA; 106733 MW; 6DED4AAB CRC32;

Query Match 26.8%; Score 302.5; DB 5; Length 964;  
Best Local Similarity 40.1%; Pred. No. 4.1e-20;  
Matches 63; Conservative 31; Mismatches 44; Indels 19; Gaps 5;

QY 32 LLLGLICIFLLKXIVRGDQPAASGDRTTTXXP-----PLPLKRRRFTPAELRRFDGVQDP 87  
Db 799 VVVVLG---FFFVLRSEQLPA-----PPKELAPLP---MSDMTVBELRYDGVKNE 845  
QY 88 RILMAINGKVFDTGKRGYGPYGVAGRDASGLATFCLDKALDEYDDLSDLTA 147  
Db 846 HILFGLNGIYDVTGKGYGPKAYGTLAGHDATRALGT--MDQNAVSEWDHTGISA 903  
QY 148 AQOETLSDESOFYFKYHHVKGKLLKEGEPTVYSDEE 184  
Db 904 DEQETANEWETQFKFKYLTGRLVKNSEKADYGNRK 940



```

KW  Heme.
SQ  SEQUENCE  877 AA;  96010 MW;  DBA0FA45 CRC32;

Query Match      8.0%;  Score 90.5;  DB 10;  Length 877;
Best Local Similarity 24.8%;  Pred. No. 3;
Matches 32;  Conservative 22;  Mismatches 44;  Indels 31;  Gaps

QY  50  QPAASGDRITTTXPPPLRLKRRDFTPAELRLRFDGVDPRLTMAINGKVFDTVKGRKFYGP 109
    QPAASGDRITTTXPPPLRLKRRDFTPAELRLRFDGVDPRLTMAINGKVFDTVKGRKFYGP 109
Db  483  EDVAAAAAIVAPPAP-AGAKSFTWAEVETHMTES--AWFVVDGKVYDATPFLKDH-P 538
    EDVAAAAAIVAPPAP-AGAKSFTWAEVETHMTES--AWFVVDGKVYDATPFLKDH-P 538
QY  110  EGPYGVF--AGRDASRGATFCIDKDEYDDLSLDTAAAOETLSDWESOFFTKYHHV 167
    EGPYGVF--AGRDASRGATFCIDKDEYDDLSLDTAAAOETLSDWESOFFTKYHHV 167
Db  539  GGADSIILVAGTDTAT-----DEFNAIHSRAKKQ-----LLEYVI 573
    GGADSIILVAGTDTAT-----DEFNAIHSRAKKQ-----LLEYVI 573
QY  168  GKLLKEGEE 176
    GKLLKEGEE 176
Db  574  GELAEQGQE 582
    GELAEQGQE 582

RESULT 7
Q22896 PRELIMINARY; PRT; 2380 AA.
ID Q22896
AC Q22896;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE SIMILAR TO TYROSINE-PROTEIN KINASE.
GN C16D9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RN Nature 368:32-38(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA GATTING S., LE T.T.;
RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64858; AAB18281.1; -.
DR HSSP; P00695; 2BQB.
DR PFAM; PF00041; fn3; 3.
DR PFAM; P00069; pkinase; 2.
SQ SEQUENCE 2380 AA; 264939 MW; 4A4DF3C6 CRC32;

Query Match      7.7%;  Score 86.5;  DB 5;  Length 2380.

```











Sequence	Strd Orig	ZScore	EScore	Len	Documentation
9b_p11:HSRPGGBIN	+ 1068.00	1980.78	2.6e-102	1941	Y12711 H.sapiens mRNA for puta
9b_p10:AF042491	+ 1002.00	1957.81	1.9e-95	1786	AF042491 Mus musculus putative
9b_cm:SSSTRMBP	+ 939.50	1852.56	3.1e-88	1893	X99714 S.scrofa mRNA for ster
9b_cm:RN05837	+ 932.00	1735.95	1.6e-88	678	AJ005837 Rattus norvegicus mRNA
9b_cm:RNU63315	+ 887.50	1642.55	1.8e-83	1885	U63315 Rattus norvegicus 25-Dx
9b_p14:AC004835	+ 535.00	940.91	2.2e-44	121803	AC004835 Homo sapiens clone
9b_p11:HSAR2030	+ 501.00	917.73	4.3e-43	1874	AJ002030 Homo sapiens mRNA for
9b_p13:AC011193	+ 340.00	571.37	8.4e-24	180439	AC011193 Homo sapiens chromo
9b_p15:AC012958	+ 307.50	522.49	4.4e-21	52149	AC012958 Drosophila melanoga
9b_p12:SPAC2B58	+ 277.00	467.96	4.8e-18	39616	AL133225 S.pombe chromosome I
9b_p12:SPAC26H5	+ 277.00	467.60	5.1e-18	41100	Z99126 S.pombe chromosome I
9b_p11:SCYPL170W	+ 262.50	468.83	4.3e-18	2209	Z73526 S.cerevisiae chromosome
9b_p11:SCLAGHXV1	+ 262.50	437.43	2.4e-16	55786	X96770 S.cerevisiae chromosome
9b_p12:AF135283	+ 244.00	442.55	1.3e-16	930	AF135283 Arabidopsis thaliana F
9b_p12:AF135284	+ 241.00	437.88	2.3e-16	847	AF135284 Arabidopsis thaliana F
9b_p12:CELM07E3	- 237.00	394.18	6.2e-14	34851	U41552 Caenorhabditis elegans
9b_p10:AC010919	- 234.50	375.99	6.4e-13	139825	AC010919 Drosophila melanoga
9b_p16:AC008336	+ 234.50	374.70	7.6e-13	159606	AC008336 Drosophila melanoga
9b_p12:AC005656	- 231.50	380.67	3.5e-13	48443	AC005656 Drosophila melanoga
9b_p12:ATAC006585	- 207.00	327.34	3.3e-10	103495	AC006585 Arabidopsis thalian
9b_p13:AC009340	+ 205.00	322.99	5.7e-10	110070	AC009340 Drosophila melanoga
9b_p14:AF173937	+ 186.50	337.15	9.3e-11	724	AF173937 Homo sapiens clone
9b_p13:AF049672	+ 169.00	307.85	4.0e-09	504	AF049672 Homo sapiens clone
9b_p12:ATT212J18	+ 166.50	253.34	4.2e-06	82912	AL132963 Arabidopsis thaliana
9b_p11:AB025603	+ 161.50	249.39	7.2e-06	48420	AB025603 Arabidopsis thaliana
9b_p15:AC010919	+ 156.00	228.76	0.0001	139825	AC010919 Drosophila melanoga
9b_p15:AC014157	- 137.50	211.03	0.0010	24429	AC014157 Drosophila melanoga
9b_p11:DMBR30C13	+ 137.50	195.74	0.0070	117674	AL122025 Drosophila melanoga
9b_p15:AC014142	- 118.00	164.58	0.3820	67936	AC014142 Drosophila melanoga
9b_p12:AC004574	- 113.00	153.20	1.65	82912	AC004574 Drosophila melanoga
9b_p16:AC008320	- 113.00	146.99	3.65	156953	AC008320 Drosophila melanoga
9b_p16:AC015817	- 105.50	125.84	54.96	325160	AC015817 Homo sapiens chromo
9b_p11:CEH38K22	+ 101.50	143.05	6.04	25610	AL024499 Caenorhabditis eleg
9b_p11:CEW03G11	+ 97.00	132.83	22.42	30762	Z67773 Caenorhabditis elegans
9b_p11:CVN129RDB	+ 92.00	157.23	0.9810	954	X56771 C.vulgaris mRNA for nitr
9b_p12:CV039930	+ 90.50	143.04	6.05	3073	U99930 Chlorella vulgaris nitr
9b_p12:AE002073	- 90.00	130.07	31.96	10596	AE002073 Deinococcus radiodur
9b_p15:AC015817	+ 89.00	96.77	2.3e+03	325160	AC015817 Homo sapiens chromo
9b_p11:SCB68	+ 89.00	115.75	200.48	38084	AL079345 Streptomyces coelic
9b_p12:AC004574	+ 88.50	107.25	596.64	82912	AC004574 Drosophila melanoga
9b_p16:AC008320	+ 88.50	101.04	1.3e+03	156953	AC008320 Drosophila melanoga
9b_p16:SG7466	+ 86.00	136.79	13.50	2454	SG7466 annexin-VI-lipid-depend

```

|||||
202 CCGCGCGCAGCGCGAGCAGCGACGACGAGCGCGCCCTCTGCCCG 251
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84
|||||
252 CCTCAAGCGCGCGACTTCACCCCGCGAGCTGGCGGCTTCGACGCG 301
84 aGlnAspProArgTleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
302 TCCAGGACCCCGCATACTATGCGCATCAACGGCAAGGTGTTCCATGTG 351
101 ThrLysGlyArgLysPheThrGlyProGluGlyProTyrGlyValPheAl 117
|||||
352 ACCAAGCGCGCAATCTACGGCGCGGCGCGGTATGGGTCTTTC 401
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
402 TGAAGAGATGCATCCAGGGCGCTTGCCACATTTTGCCTGGATAAGGA 451
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
452 CACTGAAGGATGAGTACATGACCTTCTGACCTCAGTGCCTGCCAGCAG 501
151 GluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHisVa 167
502 GAGACTCTGAGTACTGGGACTCTCAGTTCAAGTATCATCAGT 551
167 lGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspGluG 184
552 GGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTACTCAGATGAGG 601
184 luGluProLysAspGluSerSerArgLysAsn.VallLysAlaPheSerGl 200
602 AAGAACCAAAAGATGAGAGTCCCGGAAATGATTAAGCATTCAGTGG 651
200 ySerIleSer.*****TyrPheAlaLysSerPheValThrValHis** 216
552 AAGTATATCTATTTTGTATTTTGCATAATCATTTGTACAGTCCACTCT 701
217 ValPheLysThr 220
|||||
702 GTCCTTAAACA 713

```

seq\_name: gb\_ro:AF042491

```

seq_documentation_block:
LOCUS AF042491 1786 bp mRNA ROD 22-JAN-1998
DEFINITION Mus musculus putative membrane associated progesterone receptor
component mRNA, complete cds.
ACCESSION AF042491
VERSION AF042491.1 GI:2801792
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1786)
AUTHORS Kwon,S., Lunn,R.M., O'Brien,D.A., Bell,D.A. and Eddy,E.M.
TITLE The expression of a putative membrane associated progesterone
receptor component in the mouse testis and epididymis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1786)
AUTHORS Kwon,S., Lunn,R.M., O'Brien,D.A., Bell,D.A. and Eddy,E.M.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1998) LRDT, NIEHS, NIH, 111 Alexander Drive, MD
C4-04, Research Triangle Park, NC 27709, USA
FEATURES
source
location/Qualifiers
1..1786
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/tissue_type="testis"
/dev_stage="adult"

```

```

/clone_lib="Clontech ML1020b"
14..601
/notes="MAPR component"
/codon_start=1
/product="putative membrane associated progesterone
receptor component"
/protein_id="AAB97466.1"
/db_xref="GI:2801793"
/translation="MAEDVYVATGADPSELEGGGLLHEIFTSPLNLLLLGLICIFLLYK
IVRGDQPGASGNDDEPPPLPKRRDFTPAELKRDYGVQDSKILMAIKNGKVEDVTK
GRKFYEGPYGFAGRDASRGRLATCLDKALKDEYDLSLDTTPAQOETLSMDWSQF
TFKYHHVKGKLLKEGEPTVYSDDDEPRDETARKNE"
BASE COUNT 516 a 383 c 405 g 482 t
ORIGIN

```

# alignment\_scores:

Quality: 1002.00 Length: 221  
Ratio: 4.841 Gaps: 1  
Percent Similarity: 93.665 Percent Identity: 88.235

# alignment\_block:

```

US-09-203-548-1 x AF042491
Align seg 1/1 to: AF042491 from: 1 to: 1786
1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuGl 17
|||||
14 ATGGCTGCCGAGGATGCTGGCGACTGGCGCGACCGCGAGCGAGTAGA 63
17 userGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeu 34
|||||
64 GGGCGGGGGGTGCTGCACGAGATTTTACGTCTCTCAACCTGCTCC 113
34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50
|||||
114 TCCTGGGCTCTGCATCTTCTGCTCTACAGATGCTTCCGCGGGACCG 163
51 ProAlaAlaSerGlyAspArgThrThr****ProProProLeuProAr 67
|||||
164 CCGGTGCGGCGGCGCAACGACGACGAGCAACCCCGCTGCCCG 213
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84
|||||
214 CCTCAAGCGCGCGACTTCACCCCTGCGAGCTGAGGCGTTTCGATGGG 263
84 aGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
264 TCAGGACTCGGCGATTCTCATGGCCATCAACGGCAAGGTGTTCCAGGTG 313
101 ThrLysGlyArgLysPheThrGlyProGluGlyProTyrGlyValPheAl 117
|||||
314 ACCAAGCGCGCAACTTCTACGGGCTTGAGGGCCATATGGGTCTTTCG 363
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
364 CGAAGAGATGCATCCAGGGCGCTTGCCACATTTTGCCTGGACAAAGA 413
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
414 CACTGAAGGATGAGTATGACGACCTTCTGACCTCACCCCTGCACGACG 463
151 GluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHisVa 167
|||||
464 GAGACCCCTGAGTACTGGGACTCTCAGTTCAAGTATCATCAGT 513
167 lGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspGluG 184
|||||
514 GGGAAACTGCTGAAGGAAGGGGAGGAGCGCTTACTGTACTCAGATGATG 563
184 luGluProLysAspGluSerSerArgLysAsnVal.LysAlaPheSerGl 200
|||||
564 AAGAACCAAAAGACGAGACGCTCGGAAGAATGAATGAAGCATTCGGTGG 613

```

```

200 yserIleSer*****TyrPheAlaLysSerPheValThrValHis***v 217
      |||
      |||
614 ACATATCTATTTTCTATTTGCAAAATCATTTGTAAACATTCAGTCTG 663

217 alPheLysThr 220
      |||
      |||
664 TCTTTACAACA 674

seq_name: gb_om:SSSTERMBP

seq_documentation_block:
LOCUS      SSSTERMBP      1893 bp      mRNA      MAM      11-JUL-1997
DEFINITION S.scrofa mRNA for steroid membrane binding protein.
ACCESSION  X99714
VERSION    X99714.1 GI:1657408
KEYWORDS  steroid membrane binding protein.
SOURCE    pig.
ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 1893)
AUTHORS   Falkenstein,E., Meyer,C., Eisen,C., Scriba,P.C. and Wehling,M.
TITLE     Full-length cDNA sequence of a progesterone membrane-binding
            protein from porcine vascular smooth muscle cells
JOURNAL   Biochem. Biophys. Res. Commun. 229 (1), 86-89 (1996)
MEDLINE   97112407
REFERENCE  2 (bases 1 to 1893)
AUTHORS   Falkenstein,E.
TITLE     Direct Submission
JOURNAL   Submitted (01-AUG-1996) E. Falkenstein, University of Heidelberg,
            Institute of Clinical Pharmacology, Faculty of Clinical Medicine at
            Mannheim, Theodor-Kutzer-Ufer, 68135 Mannheim, FRG
COMMENT   Reference: Meyer, C.; J.Biol.Chem. 239; 726-731; 1996 (N-terminal
            protein-sequence).
FEATURES
source
1..1893
    /organism="Sus scrofa"
    /db_xref="taxon:9823"
    /tissue_type="aorta"
    /cell_type="smooth muscle cells"
    /clone_lib="Uni-ZAP XR-Vector (Fa. Stratagene)"
    /clone="PgMEF6.1"
    /codon_start=1
    /product="steroid membrane binding protein"
    /protein_id="CAA68050.1"
    /db_xref="GI:1657409"
    /db_xref="SPTREMBL:Q95250"
    /translation="MAAEDVAATGADPSELGGGLLHIFTSPNLNLLGLCIFLLYK
IVRGDPAASDSDDEPPPLRLRRDFTPAELRRFDGVQDPRIILMAINGVFDVTKG
RKFGPEGPGYGVFAGRDASRLGATFCLDKALKDEYDLDLSLTTPAQQTLDWDQSFT
FKYHVGKLLKEGEPTVYSDEEPEKDEARKND"
BASE COUNT  509 a 447 c 461 g 476 t
ORIGIN

alignment_scores:
    Quality: 999.50      Length: 222
    Ratio: 4.782         Gaps: 3
    Percent Similarity: 94.144      Percent Identity: 89.640

alignment_block:
US-09-203-548-1 x SSSTERMBP ..

Align seg 1/1 to: SSSTERMBP from: 1 to: 1893

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuL 17
|||||
42 ATGGCTGCCGAGGATGTGGCGGTACCGCGCGACCGGAGCGAGCTAGA 91
|||||
17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34

```

```

|||||
92 GGGCGGGGGCTGCTGCTGATGAGATTTTACGTCGCCGCTCAACCTGCTGC 141
|||||
34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50
|||||
142 TGCTCGGCCCTCTGCATCTCTCTACAGATCGTGC GCGGGGACGAG 191
|||||
51 ProAlaAlaSerGlyAspArgThrThr***ProProLeuProAr 67
|||||
192 CCGCGGCCAGC...GATAGCGAGCAGCAGCGCCCGCGCTGCCCG 238
|||||
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
|||||
239 CCTTAAGCGCGGACATTCACCCCTGCGAGTCGCTCGCTTCGACGGCG 288
|||||
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
289 TCCAGGACCGCGGTATCTACTATGGCCATCAACGCAAGGTGTTCCAGGTG 338
|||||
101 ThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPheAl 117
|||||
339 ACCAAAGCGCGCAAGTTCTACGGGCGCGAGGGCGGTACGGGGTCTTTGC 388
|||||
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
389 TGGAAAGAGAGCATCCAGGGGCGCTGCCACGTTTTCGCTGGATAAGGAAG 438
|||||
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
439 CCCTGAAGGACGAGTATGATGACCTTTCTGACCTCACTCCTGCCAGCAG 488
|||||
151 GluThrLeuSerAspTyrGluSerGlnPheThrPheLysTyrHisHisVa 167
|||||
489 GAGACCTGTAATGACTGGGACTCTCAGTTCACTTCAAGTACCATCAGT 538
|||||
167 lGlyLysLeuLeuLysGluGlyGluProThrValTyrSerAspGluG 184
|||||
539 GGGCAAACTGCTGAGAGGGGGAGGAGCCCGCTGCTACTCAGATGAGG 588
|||||
184 luGluProLysAspGluSerSerArgLysAsn.ValLysAlaPheSerGl 200
|||||
589 AAGAGCCCAAAGATGACAGCGCTCGGAAAATGATTAAGCGTTCGGTGG 638
|||||
200 ySerIleSer.*****TyrPheAlaLysSerPheValThrValHis*** 216
|||||
639 AAGCATATCTATTTTCTGTTATTTGCAATCATTTGTAACATTCAGTCT 688
|||||
217 ValPheLysThr 220
      |||
      |||
689 GTCTCTAAACA 700

seq_name: gb_ro:RNO5837

seq_documentation_block:
LOCUS      RNO5837      678 bp      mRNA      ROD      08-MAY-1998
DEFINITION Rattus norvegicus mRNA for putative progesterone binding protein.
ACCESSION  AJ005837
VERSION    AJ005837.1 GI:3127856
KEYWORDS  progesterone binding protein; putative.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 678)
AUTHORS   Noelte,I.
TITLE     Direct Submission
JOURNAL   Submitted (28-APR-1998) Noelte I., Biochemiezentrum Heidelberg, Inf
            328, 69120 Heidelberg, GERMANY
REFERENCE  2 (bases 1 to 678)
AUTHORS   Noelte,I., Sohn,K., Wegehingl,S. and Wieland,F.
TITLE     Rat homologue to a putative progesterone binding protein :
            molecular characterization and localization
JOURNAL   Unpublished

```

```

FEATURES
  source          Location/Qualifiers
1..678
  /organism="Rattus norvegicus"
  /strain="fisher 344"
  /db_xref="taxon:10116"
  /tissue_type="liver"
75..662
  /gene="Lewi"
75..662
  /gene="Lewi"
  /codon_start=1
  /product="putative progesterone binding protein"
  /protein_id="CAA06732.1"
  /db_xref="GI:3127857"
  /translation="MAEDVATGADPSELEGGGLQEIFTSPINLLLLGLCFLFLYK
IVRGDQPGASGNDDEPPPLRLKPRDFTPALRRYDGVQDPRILMAINGKVFDTVK
GRKFYPEGPGYGVFAGRDASRGATFCLDKALKDEYDLSLTTPAQQTINWDWSQF
TFKYHHVGVKLLKEGEPTVYSDDDEPKDEAARKSD"
misc_signal
270..287
  /gene="Lewi"
  /function="potential nuclear insertion signal"
630..653
  /gene="Lewi"
  /function="potential nuclear insertion signal"
151 a 192 c 205 g 130 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 932.00      Length: 194
  Ratio: 4.957        Gaps: 0
Percent Similarity: 96.907 Percent Identity: 90.722

alignment_block:
US-09-203-548-1 x RNO5837 ..
Align seg 1/1 to: RNO5837 from: 1 to: 678
1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17
75 ATGGCTGCCGAGGATGTGGTGGCGACTGGCGCGACCGACCGAGCTGGA 124
17 userGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34
125 GGGCGGGGGCTGCTCAAGAGATTTTCACGTGGCTCTCAACCTGCTGC 174
34 euLeuGlyLeuCysIlePheLeuLeuTyrlsIleValArgGlyAspGln 50
175 TCCTTGGCTCTGCATCTCTCTGCTCTACAAGATCGTTCCGGGGACCCAG 224
51 ProAlaAlaSerGlyAspArgThrThrThr***ProProProLeuProAr 67
225 CCCGGTCCCATGTGGGACACACGACGACGACGAGCGCCCGCTGCCCG 274
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84
275 CCTCAACCGCGTGACTTCACCCCTGCCGAACACTAGGCGATACGATGGAG 324
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
325 TCCAGGACCGCGCATCTTATGCCATCAACGCAAGGTGTTCCAGCTG 374
101 ThrLysGlyArgLysPheTyrglyProGluGlyProTyrglyValPheAl 117
375 ACCAAGCGCGCAAGTCTATGGCGGAGGGGCATACGGGGTCTTTGC 424
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
425 TGGAAGAGATGCATCCAGGGGCCCTTGCCACATTTGCTCGACAAAGAG 474
134 laLeuLysAspGluTytrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||

```

475 CACTGAAGGATGAGTATGATGACCTTTCTGACCTCACTCTCCCGACGAG 524

151 GluThrLeuSerAspTropGluSerGlnPheThrPheLysTyrHisVa 167

525 GAGACCTGATGACTGGGACTCTCAGTTCACTTCAAGTACCATCAGCT 574

167 lGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspGluG 184

575 GGGAAACTGCTGAAGGAGGGGAGGAGCCGACTGTGTACTCGGATGATG 624

184 luGluProLysAspGluSerSerArgLysAsn 194

625 AAGAACAAGAGATGAGGCTGCTCGGAAGAGT 656

seq\_name: gb\_ro:RNU63315

seq\_documentation\_block:

LOCUS RNU63315 1885 bp mRNA ROD 01-SEP-1996

DEFINITION Rattus norvegicus 25-Dx (25Dx) mRNA, complete cds.

ACCESSION U63315

VERSION U63315.1 GI:1518817

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

REFERENCE 1 (bases 1 to 1885)

AUTHORS Selmin,O., Lucier,G., Clark,G., Tritscher,A., Vanden-Heuvel,J., Gastel,J., Walker,N., Sutter,T. and Bell,D.A.

TITLE TCDD in rat liver

JOURNAL Carcinogenesis (1996) In press

REFERENCE 2 (bases 1 to 1885)

AUTHORS Bell,D.A. and Selmin,O.

TITLE Direct Submission

JOURNAL Submitted (09-JUL-1996) ICBRA, NIEHS, P.O. Box 12233, Research Triangle Park, NC 27709, USA

FEATURES

source 1..1885

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/tissue\_type="liver"

98..769

/gene="25DX"

98..769

/gene="25DX"

/note="25 Kd protein; gene is TCDD inducible similar to cytokine/growth factor/prolactin receptor superfamily"

/codon\_start=1

/product="25-Dx"

/protein\_id="AA807125.1"

/db\_xref="GI:1518818"

/translation="MAEDVATGADPSELEGGGLQEIFTSPINLLLLGLCFLFLYKIVRGDQPGASGNDDEPPPLRLKPRDFTPALRRYDGVQDPRILMAINGKVFDTVKGRKFYPEGPGYGVFAGRDASRGATFCLDKALKDEYDLSLTTPAQQTINWDWSQFSSPSTITGKLLGEEPIVYSDDDEQKRLRLGRVTEAVSGAYLFYFAKSFVTFOSVFTT"

BASE COUNT 550 a 399 c 427 g 509 t

ORIGIN

alignment\_scores:

Quality: 887.50 Length: 223

Ratio: 4.460 Gaps: 3

Percent Similarity: 89.238 Percent Identity: 81.614

alignment\_block:

US-09-203-548-1 x RNU63315 ..

Align seg 1/1 to: RNU63315 from: 1 to: 1885

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17

```

|||||
98 ATGGCTGCCAGGATGTGGTGGCGACTGGCGCGACCCAGCGAGTGGGA 147
|||||
17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34
|||||
148 GGGCGGGGGCTGCTTCACAGAGATTTCACGTGCCTCTCAACCTGCTGC 197
|||||
34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50
|||||
198 TCCTTGGCCTCTGCATCTTCTCTACAGATCGTTCCGGGGGACGAG 247
|||||
51 ProAlaLeuSerGlyAspArgThrThr***ProProProLeuProAr 67
|||||
248 CCGGTGCCAGTGGGACACAGCAGCAGCAGCGCGCGCGCGCGCTGCTG 297
|||||
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
|||||
298 CTTCAAGCGCGGTGACTTCCACCTCGCGAAGTAAAGCGATAGATGGAG 347
|||||
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
348 TCAGAGACCGCGCATCTATGGCCATCAACGGCAAGGTGTTGACGCTG 397
|||||
101 ThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPheAl 117
|||||
398 ACCAAAGGCGCGAAGTCTATGGCGCGGAGGACCATACGGGTCTTTGC 447
|||||
117 agLysArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
448 TGAAGAGATGATCCAGGGGCTTGCACATTTTGCCTGGACAAAGAAG 497
|||||
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
498 CACTGAAGGATGATGATGATGACTTTCTGACCTCACTCTGCCACGAG 547
|||||
151 GluThrLeuSerAspTrpGlu...Ser...GlnPheThrPheLysTyrHis 165
|||||
548 GAGACCTGAATGACTGGGACTCTCAGTTCAGTTCACCTTCAGTACCAT 597
|||||
166 HisVal.GlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerA 182
|||||
598 CACGTGGGAAAACACTGTTGAAGGAGCGGAGGAGCCGATTGTGTACTCG 647
|||||
182 spGluGluGluProLysAspGluSerSerArgLysAsnValLysAlaPhe 198
|||||
648 ATGATGAAGAACAAGAGATGAGCTGCTCGGAAGAGATGACTGAAGCAGC 697
|||||
199 SerGlySerIleSer***TyrPheAlaLysSerPheValThrValHi 215
|||||
698 AGTGGAGCATATCTATTTTGTATTTTGCANAATCATTTGTATACATTCCA 747
|||||
215 s***ValPheLysThr 220
|||||
748 GTCTGCTTTACAACA 763

```

seq\_name: gb\_pr4:AC004835

seq\_documentation\_block:  
LOCUS AC004835 121803 bp DNA PRI 24-MAR-1999  
DEFINITION Homo sapiens clone DJ0555N02, complete sequence.

ACCESSION AC004835  
VERSION AC004835.2 GI:4508154  
KEYWORDS HTG.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 121803)

REFERENCE  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 121803)  
AUTHORS Waterston,R.H.

TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 121803)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Mar 24, 1999 this sequence version replaced gi:3213162.

FEATURES  
source  
1..121803  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DJ0555N02"

BASE COUNT 35035 a 24906 c 25571 g 36291 t  
ORIGIN

alignment\_scores:  
Quality: 535.00 Length: 109  
Ratio: 5.144 Gaps: 0  
Percent Similarity: 95.413 Percent Identity: 95.413  
alignment\_block:  
US-09-203-548-1 x AC004835 ..

Align seg 1/1 to: AC004835 from: 1 to: 121803

1 MetaAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17  
|||||  
33922 ATGGCTCCGAGGATGTGGCGGACTGGCGCCGCCACCCAGCATCTGGA 33971

17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34  
|||||  
33972 GAGCGGGGGCTGCTGCATGAGATTTTCACGTCCGCTCAACCTGCTGC 34021

34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50  
|||||  
34022 TGCTTGGCCTCTGCATCTTCTCTACAGATCGTCCGGGGGACGAG 34071

51 ProAlaLeuSerGlyAspArgThrThr\*\*\*ProProProLeuProAr 67  
|||||  
34072 CCGCGCGCCAGCGCGGACAGCAGCAGCGCGCGCGCTCTGCCCGG 34121

67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84  
|||||  
34122 CCTCAAGCGGGCGACTTCACCCCGCGGAGCTCGGGGCTTCGACGGCG 34171

84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100  
|||||  
34172 TCCAGGACCGCGCATACTCATGGCCATCAACGGCAAGGTGTTTCGATGTG 34221

101 ThrLysGlyArgLysPheTyrGlyPro 109  
|||||  
34222 ACCAAGGCGCAATTCACGGGCCC 34248

seq\_name: gb\_pri:HSNJ2030

seq\_documentation\_block:  
LOCUS HSAJ2030 1874 bp mRNA PRI 30-DEC-1998  
DEFINITION Homo sapiens mRNA for putative progesterone binding protein.

ACCESSION AJ002030  
VERSION AJ002030.1 GI:2570006  
KEYWORDS progesterone binding protein.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1874)  
REFERENCE  
AUTHORS Gerdes,D.  
TITLE Direct Submission

159 lnPheThrPheLysTyrHisValGlyLysLeuLysGluGlu 175  
||||| :|||:::|||:::|||::||| |  
572 AGTTTAAGAAAAATATGATTATGTAGGCAGACTCTCAAAACCCAGGAGA 621

176 GluProThrValTyrSerAspGluGluGluProLysAspGluSerSerAr 192  
|||||: |||:|||||:|||||: |||:|||||:|||||:  
622 GAACCATCAGAAATATACAGATGAAGAAGATACCAAGGATCACATAAACA 671

192 g 192  
:  
572 G 672

seq\_name: gb\_htg3.AC011193

seq\_documentation\_block:  
LOCUS AC011193 180439 bp DNA HTG 01-OCT-1999  
DEFINITION Homo sapiens chromosome 17 clone 521\_p\_1 map 17, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 9 unordered pieces.  
ACCESSION AC011193  
VERSION AC011193.1 GI:6006105  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 180439)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone 521\_P\_1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 180439)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 520: contig of 520 bp in length  
\* gap of unknown length  
\* 521 6064: contig of 5544 bp in length  
\* gap of unknown length  
\* 6065 17817: contig of 11753 bp in length  
\* gap of unknown length  
\* 17818 30992: contig of 13175 bp in length  
\* gap of unknown length  
\* 30993 52243: contig of 21251 bp in length  
\* gap of unknown length  
\* 52244 77045: contig of 24802 bp in length  
\* gap of unknown length  
\* 77046 111481: contig of 34436 bp in length  
\* gap of unknown length



```

* 111482 150526: contig of 39045 bp in length
* 150527 180439: contig of unknown length
* 150527 180439: contig of 29913 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..180439
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="17"
        /map="17"
        /clone="521_P_1"
        /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 49688 a 40901 c 39134 g 49362 t 1354 others
ORIGIN

alignment_scores:
  Quality: 340.00      Length: 222
  Ratio: 2.833         Gaps: 6
  Percent Similarity: 54.054  Percent Identity: 43.243

alignment_block:
US-09-203-548-1 x AC011193 ..
Align seg 1/1 to: AC011193 from: 1 to: 180439

2 AlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuGluSe 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54217 GTGCCAAGAGATGTGGCAAGAGT..... 54243

18 rGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 35
54243 ..... 54243

35 euGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGlnPro 51
54243 ..... 54243

52 AlaAlaSerGlyAspArgThrThr***ProProLeuProArgLe 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54244 .....ACAACTTCTCC..... 54255

68 uLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyValG 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54256 .....TTNCCGAGCTGCAGCACTTCAATGGGGTTT 54286

85 InAspProArgIleLeuMetala.....IleAsnGlyLysValPheAsp 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54287 AGGACCTGCACATCTTCATGCCACACACAGGGGCAAGTGTTCANGC 54336

100 ValThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPh 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54337 GTGACCAGAGGTTCAGATTCCTCAAGGGCCAGGGCCACACACNAGGATCTT 54386

116 eAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysG 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54387 TGCCGAAGAGATACATCCAGGGGCCCTGCCACATTTGGCTGGATGAGG 54436

133 luAlaLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGln 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54437 GAGCACTGCAGGACAGCATGATGACCTTCTGACCTCCCTCGCCAG 54486

150 GlnGluThr..LeuSerAspTrpGluSerGlnPheThrPheLysTyrHisH 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54487 GAGGAGACCCNTCAGTGACTGGCCCTCTCAGTTCACATTCAAGTATTCTC 54536

166 isValGlyLysLeuLysGluGlyClnGluProThrValTyrSerAsp 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54537 ACCTGGATAAATCTGCTGAAGGATGGGAGTA.CCCACTGTGGATTTCAGT 54585

183 GluGluGluProLysAsp..GluSerSerArgLysAsnValLysAlaPheS 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54586 AGGAAGAACCAAGATGANGAATGTACAGAAAAAT..... 54622

```

```

199 erGlySerIleSer*****TyrPheAlaLysSerPheValThrValHis 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54623 .....GATTATTTACAGATCATTTGTACATTCGAG 54655

216 ***ValPheLys 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54656 TCTGTCTTAAAA 54667

seq_name: gb_htg5:AC012958

seq_documentation_block:
LOCUS AC012958 52149 bp DNA HTG 03-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC012958
VERSION AC012958.1 GI:6223365
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 52149)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CBM:10211797 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
  Source
    1..52149
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT 15435 a 10826 c 10686 g 15202 t
ORIGIN

alignment_scores:
  Quality: 307.50      Length: 503
  Ratio: 1.997         Gaps: 11
  Percent Similarity: 30.616  Percent Identity: 21.471

alignment_block:
US-09-203-548-1 x AC012958 ..
Align seg 1/1 to: AC012958 from: 1 to: 52149

10 GlyAlaAspProSer.....AspLeuGlu..... 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15298 GGTGCAGATCTCTAGTTCGGCTACGACAAGGAAATCGGCACAAATCTAAA 15347

18 .....SerGlyGlyLeuLeuHisGluIlePheThrS 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15348 CAACGATGATTCCTCATTTCTGGGCACATATCCGGGAAATCCTGTACA 15397

28 erProLeuAsnLeuLeuLeuGlyLeuCysIlePheLeuLeuTyrLys 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15398 GTCCAATGAACCTGGCCCTCTCGCCCATCATCTCTCTCTGCTATAAA 15447

45 IleValArgGlyAspGlnProAlaAlaSerGlyAspArgThrThr*** 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15448 ATCGTTCGGGATCGACCGAGTGCCATCCGTGGCGCTTGCAAGCCATC 15497

61 *ProProLeuProArgLeuLysArgAspPheThrProAlaGluL 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15498 CGAACCTGAGTTACCCAAATC...CGACGGGATTTACCGTGAAGGAGC 15544

78 euArgArgPheAspGlyValGln...AspProArgIleLeuMetalaIle 93

```



IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid c25B8 is overlapped at the 5' end by SPAC26H5.

## FEATURES

```

source
  1..39616
    /organism="Schizosaccharomyces pombe"
    /strain="972h-"
    /db_xref="taxon:4896"
    /chromosome="I"
    /clone="cosmid c25B8"
    /map="I"
  4..2703
    /note="nominal overlap with cosmid SPAC26H5 S. pombe chromosome 1"
    /label="SPAC25B8.01"
    /product="putative MAPR family steroid-binding protein"
    /protein_id="CAB61767.1"
    /db_xref="GI:6469288"
  /translation="MASTQVVFIVFLYLLITRRNRKNEKSFIASEEPKQPEWRDYT
PAELKNGKSNLSVLAIKGVYVNTMGSKFYGPQVPSAFAGHDASRLAKNSFDD
EFIPDSAEELDCSDLNDEERQALNDKMKAFPDQKYQAVGRISPREARAATISETE
EKVAHN"
  2890..3693
    /gene="SPAC25B8.02"
  2890..3693
    /gene="SPAC25B8.02"
    /note="SPAC25B8.02, len:267"
    /codon_start=1
    /label="SPAC25B8.02"
    /product="hypothetical protein"
    /protein_id="CAB61768.1"
    /db_xref="GI:6469289"
  /translation="MDVLSRVFDNEKEELDPLLNPLTASEFRAKKAELEAESIRN
GTCKTLDDLAELRRSDELEIAERWRTFLVNRQOEYEMKAAEYERYCKTLK
EMVLSHLNEKKRKIYEAKDMEDIGSESSTLLHDASQFIDRRKLHRRNAGNQNTQ
QLPSLNFDDYLLFPFTDTAVIPQSVKAVRNSVNSKPTSASLSFLSLLSWANANP
TNGRERDPRASERAERDREKAVEKGLSGATEEDIQSLQLLAKKELAKKK"
  5179..6729
    /gene="SPAC25B8.03"
  5179..6729
    /gene="SPAC25B8.03"
    /note="SPAC25B8.03, len:516,
SIMILARITY:Schizosaccharomyces pombe, O14333,
phosphatidylserine decarboxylase proenzyme 1 precursor,
(437 aa), fasta scores: opt: 762, E():0, (41.0% identity
in 493 aa)"
    /codon_start=1
    /label="SPAC25B8.03"
    /product="putative phosphatidylserine decarboxylase
proenzyme 1 precursor"
    /protein_id="CAB61769.1"
    /db_xref="GI:6469290"
  /translation="MRPQRFRRFRHPRKSVNLRGFGVGALKGVKALGMVVRYSMR
LKWNLHRRIRRRRLRGLRSLISVRPNSQWVYLLSSLPRLSLSRVWGQFNRAHLPTE
LRTFGFLYANWFGCNLSLKDPLDTHRNFDQDFCRLRPETRPVDPVPSVPSVDG
RIVCGVYVNDNRKIQHVKGLSLEALLGSISSNPLVNFNFEIDPDLIQKHEQAEQ
HSISLNSNRYNRKADSAAVVDEHDEEALLCAFTDHPHFYLDNSRNSLNYFCPSAF
EDISNVSRS CGKRISPSNFDLNNLGDDDLRSESSDSESAPASILEHEPTNWDMM
VQEAQVTDIDSLPWHNRPGNKLFFSVIYLAGDPYHRFSPADWYERSRHFGSELFS
VSPETARRLNHLNVLNERNVALLGRVYEHGFMSPMTPVGTATNGSVTIVNCDPTLSNRLVL
RKSLCTFQEAAYKNASPVLDGMPVSRGEQVGQFQLGSIWLVFEAPADEFTYQOQ
YVRVGEAL"
  5744..5872
    /gene="SPAC25B8.03"
    /note="possible inframe intron"
    complement(7203..8339)
    /gene="SPAC25B8.04c"
    complement(7203..8339)
    /gene="SPAC25B8.04c"
    /note="SPAC25B8.04c, len:378, SIMILARITY:Saccharomyces
cerevisiae, MS51_YEAST, mss51 protein., (436 aa), fasta
scores: opt: 641, E():8.9e-33, (36.7% identity in 398 aa)"
    /codon_start=1
    /label="SPAC25B8.04c"
    /product="similar to S. cerevisiae mitochondrial pre-mRNA
maturation protein Mss51p"
    /protein_id="CAB61770.1"
    /db_xref="GI:6469291"
  /translation="MSFLRNLFKKSPTHELPHLSRSPLTALRRSEHIKQVYRCPI
SGKSVYECPESGFTPCNTRHWEQKHQSLIPKLRQINEDYHDLARNPLPELKL
PGMEDEYVSLSDSFYTRDFPKFSQSRARHTISLTTPMSIGAILHKNSPYNL
KNGLTAPQGLQSLTALRYMLHRPLSAQSTDPRTTRIFVLGATKESCPSPISWLOGLNPL
FGRQLFQLHFIGPEVVPVSKQNPSPSLSHHQDYHNLHRVGATEFEPDPYDTFFL
PINKPLPIYSSSWITPDLHDLSTRCSWLTSPSSORTKDLVNLNVLKDSIEPLLL
PVTNFKASLGWSVDDSNLSEVYHANOVEGFRALYVNVQNSKE"
  9151..10503
    /gene="SPAC25B8.05"
  9151..10503
    /gene="SPAC25B8.05"
    /note="SPAC25B8.05, len:450, SIMILARITY:Saccharomyces
cerevisiae, PUS3_YEAST, pseudouridylate synthase 3, (442
aa), fasta scores: opt: 1064, E():0, (43.8% identity in
454 aa)"
    /codon_start=1
    /label="SPAC25B8.05"
    /product="probable pseudouridylate synthase"
    /protein_id="CAB61771.1"
    /db_xref="GI:6469292"
  /translation="MPASKVAGWSRDQLISRLLELDGKLQKNTNSLPNKEIECKNEVI
KNKNLDNGNEIEPQHSQELVQLHALQDASSLPKLPVRPQLEKTVVAFKFAFCG
WNYNGLAYOLEPTPLPTVEGKVFALLRAHLITDPSCSFSGCRGTDKGVSAWGQVIS
LWVHRSERRPIYCDVNLRLPADITGYAFPEGFDAFSCQKRYLFRTKDYPG
GLSIDIRMNAAALYLGEHDFNFKIDASKQITNVHRRILSKVICIDPSTGLYAFD
LOGTAFLHQVRCMAAILFLIGOKLEPASLINDLDIKVPTKPIYDMASEVPLIYD
CFDNIETWLTPTDSPTTAPRIAKHYETIYHWHSLRIREQTASFLMDLTADHNVKRYG
KSEESSSTWNVGEGLLKRTKYPILPSRQESVEVNVNERYKRRKGLKDSQVKSNIK"
  9505..10095
    /gene="SPAC25B8.05"
    /note="Match to PF01416 Pseudou synth_1, trna
pseudouridine synthase Score 233.73"
    complement(10629..12050)
    /gene="SPAC25B8.06c"
    complement(join(10629..10950,11008..12050))
    /gene="SPAC25B8.06c"
    /note="SPAC25B8.06c, len:454, SIMILARITY:Saccharomyces
cerevisiae, YHH1_YEAST, putative seryl-trna synthetase
yhr011w, (446 aa), fasta scores: opt: 935, E():0, (40.1%
identity in 436 aa)"
    /codon_start=1
    /label="SPAC25B8.06c"
    /product="probable seryl-trna synthetase"
    /protein_id="CAB61772.1"
    /db_xref="GI:6469293"
  /translation="MSQLYLRFTSTHTSVLRSSWQAILNKYVIYQNAVORNCINR
NQAAETVPKIRSLIDKESLKNFEPLLSLKKETLTOIERCSDNERGKLVNEAKG
LKKTEEYNNKISKYNDLYQYCLAVPNTLTPTVPGPDKAVVQVQKISPLVKKTGS
LKDHQIQIANEGINLEDAAGSHGSCYTTGDIALLEMAITNYAMDFAISRGWCPVIPP
TVTRDIALACGFGPDEAQGIYELDSYSPVSSPKQCLGTABISLAALGFKKFT
NNFTERKVVGVSVRAVREARGAKENRGLRYLHEFTKVLFATWHTPSRSSEMEINIV

```

FQKEFVETLKIPIARILNMPTAELGSSAQYDIEMWPARQSYGEITSASNCLEYQAR  
RLTRYRNDKSGVHFLNCTAAAIPLRIITALENHQOEDGTVKVPETLVPIHKEYL  
FKAK"

misc\_feature complement(join(10659..10950,11008..11777))

/gene="SPAC25B8.06c" /note="Match to PF00587 tRNA-synt\_2b, tRNA synthetases

class II (G, H, P and S) Score 67.23"

misc\_feature complement(10951..10963)

/gene="SPAC25B8.06c"

/note="ctaacaatatag, splice branch and acceptor"

misc\_feature complement(11002..11007)

/gene="SPAC25B8.06c"

/note="gtacat, splice donor sequence"

gene complement(13662..14003)

alignment\_scores: Quality: 277.00 Length: 121

Ratio: 3.044 Gaps: 2

Percent Similarity: 75.207 Percent Identity: 45.455

alignment\_block:

US-09-203-548-1 x SPAC25B8 ..

Align seg 1/1 to: SPAC25B8 from: 1 to: 39616

71 ArgAspPheThrProAlaGluLeuArgPheAspGlyValGlnAspPr 87

1944 AGAGATTATATCCAGCAACTGAGGATATATGTCCTCAAAATTC 1993

87 oArgIleLeuMetAlaIleAsnGlyLysValPheAspValThrLysGlyA 104

1994 ATTGGTTTCTTGCTATCAAGGAACTGTGTACAACTGTACAAATGGAT 2043

104 rGlyPheThrGlyProGluGlyProTyrGlyValPheAlaGlyArgAsp 120

2044 CAAATTTTATGTCACAAAGCCCTTATAGTCTTTGCGGTCATGAT 2093

121 AlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluAlaLeuLysAs 137

2094 GCCAGTCGTGGTTAGCAAAAATAGTTTGGAGCATGATTTATCTCTGA 2143

137 p.....GluTyrAspLeuSerAspLeuThrAlaAlaGlnG 150

2144 TTCAGTCGAGAGAAATGTCAGTATTCAGTATTTAAACGATGAGGAGA 2193

150 InGluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHis 166

2194 GGCAGGCTCTAAACGATGCGAAAGCATTTTTCACCAAAAATATCAAGCT 2243

167 ValGlyLysLeuLeu.....LysGluGlyGluGluProThrValTyrSe 181

2244 GTTGCAGCTGATAGTCCCGAGAGGCCAGACGACGACTACTATTTC 2293

181 rAspGluGluGlu 185

2294 AGAACCGAGAA 2306

seq\_name: gb\_pl2.SPAC26H5

seq\_documentation\_block:

LOCUS SPAC26H5 41100 bp DNA PLN 03-SEP-1999

DEFINITION S.pombe chromosome I cosmid c26H5.

ACCESSION 299126

VERSION 299126.1 GI:2398810

KEYWORDS beta transducin; chromatin assembly factor; dna-directed rna polymerase mitochondrial; glucan 1,3-beta-glucosidase precursor; helicase; initiation factor elf-5a; LTR; mannitoltransferase; oxidoreductase; steroid binding; ttf51. fission yeast.

SOURCE Schizosaccharomyces pombe

ORGANISM Eukaryota; Fungi; Ascomycota; Archiascomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomyces.

REFERENCE

AUTHORS Oliver,K. and Harris,D.

JOURNAL Unpublished

REFERENCE

AUTHORS Barrell,B.G., Rajandream,M.A. and Wood,V.

TITLE Direct Submission

JOURNAL Submitted (11-SEP-1997) Schizosaccharomyces pombe chromosome I

sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk

COMMENT

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, http://www.sanger.ac.uk/Projects/S\_pombe/)  
protein coding regions (CDS) have been predicted with the help of computer analysis using the genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: it is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: this sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid c26H5 is overlapped at the 5' end by cosmid c23A1.

Location/Qualifiers

source

1..41100

/organism="Schizosaccharomyces pombe"

/strain="972h-"

/db\_xref="taxon:4896"

/chromosome="I"

/clone="cosmid c26H5"

/map="IR"

complement(1..1929)

/gene="SPAC26H5.01c"

/note="nominal overlap with cosmid c23A1 em.AL021813"

complement(1..1929)

/partial

/gene="SPAC26H5.01c"

/note="SPAC26H5.01c, len:643, SIMILARITY: Saccharomyces

cere visiae, Q05549, similar to several dna helicases.,

(1077 aa ), fasta scores: opt: 1282, E(): 0, (38.4%

identity in 633 a a)"

/codon\_start=1

/product="putative dna helicases."

/label="SPAC26H5.01c

/protein\_id="CAB16187.1"

/db\_xref="GI:2656007"

/db\_xref="SPTREMBL:O13983"

/translation="MSQTPIKKEESNDQDDKFEFKYINEGKLPKADNPKKPKQLGT IQANQIPISFDNLFNFKVINTTYFLYLNSTITFPLNNSVKQSLKELTIGDL SOLREICPOIELNYKSLASLAFINKVNTDLNPELYTGSTVSQSEYVLIVLELT QERSKRRREGPTMKANIQRKLDFFNNLKAIELRQKFLQGIKEIKKQLTELDP TSSGKSLTYQIPIQLSLEDNQSTAFVFTFKSLAQDQKSLDILSYMPTLKNRIVD TFGDPTLESRESIIRANSIIFTNPDMLHOTILPNANRWYFFKNLFLVDEAHVYN FIGFVHVFLVLRMRRIAEEYNSQYRFVSCSATIEDPLQHKMKIFGVDNKLINYS SPGSKAFVWNNPYDPKHPDDGKSAISAKLIRFAEKRVRTIVFCRVKRTICES LMRLVROELKTKQKDLLSKIQSYRAGTYQVERKRIESEMFNGKLYGIATNALELGI DIGSLDAVITIGFPYLSLN"

complement(2096..3721)

/gene="SPAC26H5.02c"

complement(join(2096..2948,2993..3010,3056..3086, 3130..3721))



```

40591 GGCAGGCTCTAAACGATTGGAAGCATTTTTCACCAAAATATCAAGCT 40640
      :::::::::::::::::::: ||| |||||
167 ValGluLysLeu.....LysGluGlyGluGluProThrValTyrSe 181
      |||||:::|||||:::
40641 GTTGCCAGCTGATAGTCCAGAGAGGCCAGAGCAGCAGCTACTATTTC 40690
      ::::::::::::::::::::
181 rAspGluGluGlu 185
      ::|:: |||||
40691 AGAAACCGAAGAA 40703
      ::|:: |||||

seq_name: gb_p11:SCYPL170W
seq_documentation_block:
LOCUS SCYPL170W 2209 bp DNA PLN 07-AUG-1997
DEFINITION S.cerevisiae chromosome XVI reading frame ORF YPL170W.
ACCESSION Z73526 U00094
VERSION Z73526.1 GI:1370358
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS Benes,V., Rechmann,S., Nentwich,U., Voss,H. and Ansoerge,W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2209)
AUTHORS Purnelle,B., Coster,F. and Goffeau,A.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2209)
AUTHORS MIPS.
JOURNAL Direct Submission
TITLE Submitted (28-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome XVI sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
source
1..2209
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XVI"
complement(1..542)
/gene="OYE3"
complement(<1..542)
/gene="OYE3"
/note="ORF YPL171c"
/codon_start=1
/protein_id="CAA97877.1"
/db_xref="GI:2326834"
/db_xref="SWISS-PROT:P41816"
/translation="MPFVKGFEPISLSDTNLFEPKIGNTQLAHRAVNPPLTRMRATH
PGNIPNKWAAVYIGQRAQRPMTIITEGTFISPGAGGYDNPAGIWSDEQVAEKNI
LAITHDCQSAFWOLMSLGSWASFPDLARDGLRYDCASDRVMYMNATLQEKARDANNLEH
SLTKDDIKQYIKDYIHAAKNS"
CDS
1485..1943
/note="ORF YPL170W"
/codon_start=1
/protein_id="CAA97876.1"
/db_xref="GI:1370359"
/db_xref="SPTREMBL:Q12091"
/translation="MSPIKLLFGVKTSDPTGLTGNGASNTNDSNKGSEPPVVGNF
FPRTLKFNHGDDEKFIALRGKYVDCTRGQYFGPSGPTNFAGHASRGLALNSFD
LDVKNWDQIDPDLDTKEQIDALDQWHEFNKFCIGTLIPEPGNV"
BASE COUNT 654 a 450 c 476 g 629 t
ORIGIN

alignment_scores:
Quality: 262.50 Length: 170
Ratio: 2.625 Gaps: 4
Percent Similarity: 58.824 Percent Identity: 35.882

alignment_block:
US-09-203-548-1 x SCYPL170W ..
Align seg 1/1 to: SCYPL170W from: 1 to: 2209

7 ValAlaThrGlyAlaAspProSerAspLeuGluSerGlyGlyLeuLeuH1 23
|||::: |||||::: |||::: |||::: |||::: |||:::
1518 GTTAAACAAGTGAAGGATCCACCGGGCTCACAGGTACGGGCGCTCAAA 1567
23 sGluIlePheThrSerProLeuAsnLeuLeuLeuGlyLeuCysIleP 40
: |||::: |||
1568 C.....ACAAACGATTCTAAT..... 1583
40 heLeuLeuTyrLysIleValArgGlyAspGlnProAlaAlaSerGlyAsp 56
: |||::: |||::: |||::: |||::: |||::: |||:::
1584 .....AAGGTAGTAGCAGCGGTAGTAGCGGT... 1610
57 ArgThrThrThr***ProProProLeuProArgLeuLysArgArgAspPh 73
: |||::: |||
1611 .....AATT 1615

73 eThrProAlaGluLeuArgArgPheAspGlyValGlnAspProArgIleL 90
| ||| ||| ::|||::||| ::|||:: |||::: |||:::
1616 CTTTCTAGGACGCTTCCAAATTTAACGCCGCCAGCATGAAAAATAT 1665
90 euMetAlaIleAsnGlyLysValPheAspValThrLysGlyArgLysPhe 106
::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1666 TTATTGCTATTAGGGCAAGTATACCACTTGCACAGAGGGGCGAGTTT 1715
107 TyrGlyProGluGlyProTyrGlyValPheAlaGlyArgAspAlaSerAr 123
||| |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1716 TAGCGTCCAAAGCGGCATACACTAATTTCAGGCCATGATCGTCGCG 1765
123 gGlyLeuAlaThrPheCysLeuAspLysGluAlaLeuLysAsp..... 137
||| |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1766 TGGCTCTGATGAACCTCTCGACTCGACGCTTATTAAGATTGGGATC 1815
138 ..GluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153
||| |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1816 AGCTATATCGATCCCTAGATGATCTGACAAAGACAGATTGACGCGCTG 1865
154 SerAspTrpGluSerGlnPheThrPheLysTyrHisHisValGlyLysLe 170
::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1866 GATGAGTGGCAAGAGCATTTTGAGATAAGTACCCCATGCGATTGGTACTCT 1915

170 uLeuLysGlu 173
::: |||
1916 GATTCGGAG 1925

seq_name: gb_p11:SCLACHXVI
seq_documentation_block:
LOCUS SCLACHXVI 55786 bp DNA PLN 03-DEC-1996
DEFINITION S.cerevisiae chromosome XVI, left arm DNA.
ACCESSION X96770
VERSION X96770.1 GI:1403537
KEYWORDS BEM4 gene; CDC60 gene; KES1 gene; KIP2 gene; OYE3 gene; PALL1 gene;
PEP4 gene; PXA1 gene; REV3 gene; ribosomal protein L37a; RPL37A
gene; SNR17B gene; spk1 gene; SVS1 gene; u3 small nuclear rna.
baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 55786)
AUTHORS Purnelle,B., Coster,F. and Goffeau,A.
TITLE The sequence of 55 kb on the left arm of yeast chromosome XVI
identifies a small nuclear RNA, a new putative protein kinase and
two new putative regulators
JOURNAL Yeast 12 (14), 1483-1492 (1996)
MEDLINE 9710377
REFERENCE 2 (bases 1 to 55786)
AUTHORS Purnelle,B.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1996) B. Purnelle, Unite de Biochimie

```

Physiologique, Université Catholique de Louvain, Place Croix du Sud  
2/20, 1348 Louvain-la-Neuve, BELGIUM  
COMMENT Overlapping sequences: L29279, M29683, D50278, L27816, X62878,  
Z11963, M13358, M55623, L38491, U17065, U03913, X57969, X05498.  
FEATURES  
source  
1. .55786  
Location/Qualifiers

/organism="Saccharomyces cerevisiae"  
/strain="AB972"  
/db\_xref="taxon:4932"  
/chromosome="XVI"  
/map="left arm"

/clone\_lib="pHC79 or Supercos derivative cosmids (L. Riles  
St Louis, Missouri)"

/clones="cos9574"

/clones="cos8484"

complement(1. .542)

/gene="OYE3"

complement(<1. .542)

/gene="OYE3"

/note="L29279"

/codon\_start=1

/db\_xref="SGD:S0006092"

/product="P2591 protein"

/protein\_id="CAA65573.1"

/db\_xref="GI:1403538"

/db\_xref="SWISS-PROT:P41816"

/translation="MPVKGEFSLRDTNLEPIKIGTQLAHRAVMPPTMRATH  
PGNIPKEAWAYIGQRAQPGTMIITETGTFISQAGGYDNPAGTWSDEQVAEKNI  
LAIHDCSFVAVQWLSGWSGFPDYLARDGLRYDCASDRYMNATLQEKAKDANLEH  
SLTKDDIKQYIKDYIAHAKNS"  
1485. .1943

/note="putative"

/codon\_start=1

/product="P2515 protein"

/protein\_id="CAA65551.1"

/db\_xref="GI:1403539"

/db\_xref="SPTREMBL:Q12091"

/translation="MSPIKLLFGVKTSDEPTGLTNGASNTDNSKGSPPVAGNF  
FPYLSFNGHDEKIFIAIKGVYDCTRGQFVGPSTYFNGADSRGLALSFD  
LDVTKDWDQPTDLDLTKEQIDALDEWQEHFNKYPCIGTLIPEPGVNV"  
complement(2210. .4009)

/note="putative"

/codon\_start=1

/product="P2520 protein"

/protein\_id="CAA65552.1"

/db\_xref="GI:1403540"

/db\_xref="SPTREMBL:Q99257"

/translation="MSGFHNVTNNMAQQQNRKIKSVRNQONATNDLINFIS  
NARVAVDAHVEGLVIGYNSKAEASLKWNGVRFAGSNLKFELLDNNGASAGTSD  
TISFLRGVLLKRYDPQTKLNLGALHSDPELIQGVFSSISTQKMFPMKKLASEK  
SLIVSVNLADNQLKDISAISTLAQTFPNLKLANNQIIFRFSLEYKMKFKDLRE  
LMTNNPTDKLRTMLNRPFLKVLNDNVIRDEQLOTVYSLPMKIQOFPENDAL  
LGQSDTFATNFLNDRNREQLNLXSPQSFVSDVSTIPPSVTVDOTPAFCY  
MSSSRNISKVSSEKSIQORLSIGQESINSIKTLPTKHHLEQNEYSMTISYPI  
NGVITLHGFEETGKLESTNKKTKNNYQNNRRYNGHNSNNKLSKSFSDTW  
IVPMNNVSIADSLTVRAYSTGAWKTASIAIQPQQQASVLPQVAMNPNITPPQ  
PQSPVPGGSIIPGAPQAMVMAPTLQLPDQVQSLNVPQLELLNKLKLTALNAEY  
FMLAEQSNWNYEVAIKGFQSSMNGIPREAFVQF"  
4388. .5680

/note="putative"

/codon\_start=1

/product="P2525 protein"

/protein\_id="CAA65553.1"

/db\_xref="GI:1403541"

/db\_xref="SPTREMBL:Q12467"

/translation="MTVLTSALKMKKCLAFNMGMNCVTVSHARSGAKFGGRNVF  
NIFDSKTPDSVRKAFKNTIYOSAMGKTKFSAMEINLITSLVRGYGEGKKNAINP  
LQTNVQIKLLTHRLTKDQILEGMNLAAGPVNVAIPRDTIPDEKKKVELRNKAE  
NMDLHPSRKMKHIKELLHSLNDMCNDEEVYOKISLYLOKNEESTSVGASQONHVKDID  
INSLKRVQLNIEKKARKQSAIDKQKNQARIYQNTQSFSEIVPLSAGNILEFKREP  
LWKRLQNGISVFLGNGGGKSKTKKVLQGNLILLHLENNKMTLSNFDHVSFNI  
NFTDLFGVNASGPPDRVLNEINEILKWKVCVGNLYDNNKIVVVFQSNPLLEDTKI

gene

CDS

PKQSTNSKRFLISLSALASFAFFAYRYRLSQRQESKK"

complement(5764. .10278)

/gene="REV3 (also called PS01)"

/db\_xref="SGD:S0006088"

complement(5764. .10278)

/gene="REV3 (also called PS01)"

/note="M29683"

/codon\_start=1

/db\_xref="SGD:S0006088"

/product="P2535 protein"

/protein\_id="CAA65554.1"

/db\_xref="GI:1403542"

/db\_xref="SWISS-PROT:P14284"

/translation="MSRESNDTIQSDTVRESSKSDYFRIQLNODYMSKPTFLDP  
GESLPLNQFSQVPIRVFALPTGQVCHVGLIPYMEIKYDGOITDSTLRHQRCA  
QVHKLTVKIRASFRKKDKDLGKGLNLFVADVSVVGIPGYGVHGNLRYK  
ISLNPSCLSRISLIDGKIFGKFEIYESHVYLLQWADFNGCWINVDRCYF  
RSPVLNSLIDIKLATINDQLLDRFCFKNVLSRRDFPRVGNGLIETIDILPQIK  
NREKLQHRDIHHDLEKDISDIPVKYVSARDMINELTWORELSLKYKEPPE  
KRVSHQWSSSEFEAFYKKAQHTSTEDGOIPNFENIDKNQKFSAINPYEALPO  
LWPLRPQIEINNNMQDKNDQVNASFTEYICGVNENEGVSGNSKRSYSWLP  
SIASPKDSTILLDQTKYHNTINFSDCMTQNMASKRLRSVSANKTSLSRKRRK  
VNAAGLRXKRAVYCEPFGYQDILNKLDEGFPKIDYKDPFNSPVLNKPAY  
GKRFEISSTHVSTRIPVQGGTGVYNNKPTDFMSSWYALKPPTYDAVQWYKVP  
SMGNKTESQISMTHTPHSKFLYKASDVSGKOKRKKSSVHDSLTITLHANTRS  
IPDPAIDVSMIWCLEETFPDLIDIAEGIMVHKASEDSTFFPKIQHCINEIPW  
FYSEFMEALDVLDDPILSGFEIHNFSWGIIERCOKIHOFDIVRELARVKC  
QTKLSTDTGYAHSGMIGRMINIRALRSVDNLTQYTESAAFILHKLPHF  
SPESITNNMNAKSTTELKTVNLWSRAQINQLKQDYIARNTEQARLIDPFS  
VYRSGQKVESFLIRICKSEFILLSPOKQVOKKALECVLWMEPSAFYKSLPI  
VLDFOSLPSIMGYNCYSTMIGVREINLTENNLSKVSFSLPNILALLKNDVTIA  
PNGVYAKTSYRKSLSKMLDILQVRYMKTMEIGDNTTLKRLNKKQALAKLL  
ANVTYTSASFSGRMPDLSADISVQTRGLEKADIDIEKDETNAKVVYDGTDSL  
FVYLPKTAIEAFISCHAMAEVQNNPKPIFKFVKYVHPSILSKRYGVFSYSP  
SQTLPIDAKIETVTRDGIQAQKRIEIKRLLFKTQKLSIKKYLQNEFKIQIGK  
VSAODFCFAKEVKLAGSEKAPAGAVVVKRINEDHRAEPQYKRIPLYLVKQGG  
OLLRERCYSPFEELGENLELDSEYINKLIPDLRLNLIGINVNAQELVSKR  
ASTTTKVENTRVTGTSATCCNGEELNICSQJLCCDCEKRSSTTTLFLIKKLAKQ  
KEQYTLKTVCTCSYRTSDAGIENDHASKNSYDCPVFSYRVKAEYLRDNQSVOR  
EEALISLNDM"

/note="putative"

/codon\_start=1

/product="P2540 protein"

/protein\_id="CAA65555.1"

/db\_xref="GI:1403543"

/db\_xref="SPTREMBL:Q12092"

/translation="MIMNSTVVIKVKRRPQGFDPDPKFEWNGTKERQLTWYSN  
LNSQDQDMQNLKIFETPEFLKRTYKLAHLELLQLQLEKRLDEKYSNDQVN  
EGMSDLIIHKYPTLQNDNLNVASAPLTERQDSEEVETEVTNEALQHLQTSKILNIH  
KRTSDSENKPNKLDKDGINKEMEGCGSSDDLSLSVSKSALEALMDRLQF"

CDS

complement(11127. .12248)

/note="putative"

/codon\_start=1

/product="P2545 protein"

/protein\_id="CAA65556.1"

/db\_xref="GI:1403544"

/db\_xref="SPTREMBL:Q12529"

/translation="WTIDGDVHEISPFQVROTQWGGACFNGNTPKGTTVLOVSNF  
TGTSTSYFRKEVCHNCFAYANAKTKYKLYNDYLDRLVCNHAQINPKRFLGAGWF  
CSEHRTSYLQIPNTIELTEYEILLHFPFSLMKRYNTSEQEEKNSLISENTIQS  
SWDEIESKWIPRNNKSAKRINQLPPTCEDEYCGTFRVCELSFKLKYMDPQCITVRA  
FNMQSNELSKISKFPVLLHFKLVQFVLYILLPNHRLMSIPLLRHLTLGTGYGNF  
GLWQGEASDSREYGVYVFEASVENHSCNPNITKYRKGNSMLFTMNRDIDKDEQIC  
IDYSGVLDLPTVKRAFLADSWFFDCAERCACKSELQSVH"

CDS

complement(12521. .14668)

/note="putative"

/codon\_start=1

/product="P2550 protein"

/protein\_id="CAA65557.1"

/db\_xref="GI:1403545"

/db\_xref="SPTREMBL:Q12083"

/translation="MSQHIRKLDSDVSRKLSQACTVSLASAVREIVQNSVDAHATTI

CDS

DVMDIPNLSFAVYDDIGLIGTRSDNLILATONTYTSKIRKMDLVTKTKYGYRGDALYS  
 IENVSLNFCVKSKNDYNAMRRFPKSPKSVMLSENTILPIDPFWKICPWSRTKSGTVVI  
 VEDMLNLYFRRRIILKEEPFKFTNTIKADMLQILVHPMISLNVQVTDKRLNTLEVL  
 FRSKNTGLELTHQOOSVLRNVEGAILPPDMLKVKSLKFNEXQIEGIIISKMPVGLKO  
 LOFIYINGRYADSAGQGVYDSLFOQDFGEKMSLLKTKSVGKPRSRHPVFIIDVRC  
 PQIIDLQOPAKIKVPKSHIRTIETLIVKIRSFQGLTPDKSDSSFEIVNCQS  
 KATLPDSRIQISKRNOVLSNKMKIARINSYIGKPAVNGCRINNSYNEKIKNIRID  
 GOKSRINKLSSRPISGFTEDYSIKTITDFSISRSVLAKYEIVNQVDFKILIRIC  
 LQOSIRNCLLVLDHACDERIRLEELFYSLTVEVGTGFVARDLKDCCIEVDRETA  
 DLFKYQSEFKWIGYETIEGTMETSLLEIKTLPMLTSTKNGDKDYLKMWILLOHAE  
 DLKDFKILPMDLSHFENYTSVDKLYWKKYSSCVPTVFHEILNSKACRSAYVFGDELTR  
 QECIILISKLRSCHNFECAHGRPSWVPIAELK"  
 complement(15089..15871)  
 /gene="SVS1"  
 /db\_xref="SGD:S0006084"  
 complement(15089..15871)  
 /gene="SVS1"  
 /note="SVS1"  
 /codon\_start=1  
 /db\_xref="SGD:S0006084"  
 /product="p2554 protein"  
 /protein\_id="CAA65558.1"  
 /db\_xref="GI:1403546"  
 /db\_xref="SWISS-PROT:Q12254"  
 /translation="MIFKILCSLLVTSNFSALYVNETTSYPTTKLTPTYSVSPQ  
 ETLTYSDETTFYITSTFYSTYFWFTSQSAALISTPASTPASTPSLTSTNXYTT  
 TYSDDTTVTSTLTSTYIITLSTESANERAOISTVTEASTVTSGSTYTSTLTST  
 LVTYVNSOASNTIANTAGDAASNDVALEKLYSAEHSQSMQITTSNDEYCSASIKY  
 VVTAAATVEVTTTAEVYKVTITADASNVTGSANNGTHI"  
 complement(16377..17198)  
 /note="putative"  
 /codon\_start=1

## alignment\_scores:

Quality: 262.50 Length: 170  
 Ratio: 2.625 Gaps: 4  
 Percent Similarity: 58.824 Percent Identity: 35.882

## alignment\_block:

US-09-203-548-1 x SCLACHXVI ..

Align seg 1/1 to: SCLACHXVI from: 1 to: 55786

7 ValAlaThrGlyAlaAspProSerAspLeuGluSerGlyGlyLeuLeuHi 23  
 1518 GTTAAACAACTGAGGATCCAAACCGGCTCAGAGTAACGGCGCTCAA 1567  
 23 sGluIlePheThrSerProLeuAsnLeuLeuLeuLeuGlyLeuCysIleP 40  
 1568 C.....ACAAACGATTCTAAT..... 1583  
 40 heLeuLeuTyrlLysIleValArgGlyAspGlnProAlaAlaSerGlyasp 56  
 1584 .....AAAGGTAGTAGAACCAGCGGTAGTAGCGGGT... 1610  
 57 ArgThrThrThr\*\*\*ProProLeuProArgGlyLysArgArgAspPh 73  
 1611 .....AATT 1615  
 73 eThrProAlaGluLeuArgArgPheAspGlyValGlnAspProArgIleL 90  
 1616 CTTTCCTAGGACGCTTCCAAATTAAGCGCCACGACGATGAATAATAT 1665  
 90 euMetAlaIleAsnGlyLysValPheAspValThrLysGlyArgLysPhe 106  
 1666 TTATTGCTATTAGGGCAAGTAAGTACGACTGCACAAGAGGAGGAGGAGTT 1715  
 107 TyrGlyProGluGlyProTyrlGlyValPheAlaGlyArgAspAlaSerAr 123  
 1716 TACGGTCCAAGCGGCGCCATACACTTAACCTTTCAGGCCATGATCGTCGCG 1765  
 123 gGlyLeuAlaThrPheCysLeuAspLysGluAlaLeuLysAsp..... 137  
 ||||||| :|||| :|||| :|||| :||||

1766 TGGTCTTGCACTGAACCTCCTTGATCTGGACGCTATTAAAGATTGGGATC 1815  
 138 ..GluTyAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153  
 1816 ACCCTATGATCCCTTAGATGATCTGACAAAAGAACATGATGACGCACTG 1865  
 154 SerAspTrpGluSerGlnPheThrPheLysTyrlHisHisValGlyLysLe 170  
 1866 GATGAGTGCAAGACCAITTTGAGAATAAGTACCCATGATGATGATGATCT 1915  
 170 uLeuLysGlu 173  
 1916 GATCCGGAG 1925  
 seq\_name: gb\_pl2:AF153283  
 seq\_documentation\_block:  
 LOCUS AF153283 930 bp mRNA PLN 02-JUN-1999  
 DEFINITION Arabidopsis thaliana putative progesterone-binding protein homolog  
 (Atmp2) mRNA, complete cds.  
 ACCESSION AF153283  
 VERSION AF153283.1 GI:4960153  
 KEYWORDS  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 REFERENCE 1 (bases 1 to 930)  
 AUTHORS Choi,J.H., Choi,H. and Gray,P.  
 TITLE Plant homologues of mammalian putative progesterone-binding  
 membrane proteins  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 930)  
 AUTHORS Choi,J.H., Choi,H. and Gray,P.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-MAY-1999), School of Biology, Georgia Institute of  
 Technology, MC 0230, Atlanta, GA 30332-0230, USA  
 FEATURES  
 source  
 1..930  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /db\_xref="dbEST:U14095"  
 gene  
 1..930  
 /gene="Atmp2"  
 CDS  
 12..773  
 /gene="Atmp2"  
 /codon\_start=1  
 /product="putative progesterone-binding protein homolog"  
 /protein\_id="AAD34615.1"  
 /db\_xref="GI:4960154"  
 /translation="MVQIWEITLKITATYGLSPAFFTVLALAFAYOVWVGFFVS  
 PEVRRPSLEVQPSPELPVOLGEITTEELKLYGSDSKKPLMLAKGQIYDVQSOS  
 RMEYGGPYALEFAGDASALAKMSFEQDLTGDISGLGAFLEALQDWEIFKMSKY  
 VKVGTOKKDGKESSEPEAKTAEGLSTNTGEEASAITMMKULEAQARKSRKPR  
 RRKMLQLMMMLQSRKVNVRGTVSLEKLVRFVR"  
 BASE COUNT 281 a 183 c 218 g 248 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 244.00 Length: 172  
 Ratio: 2.324 Gaps: 4  
 Percent Similarity: 61.047 Percent Identity: 34.884  
 alignment\_block:  
 US-09-203-548-1 x AF153283 ..  
 Align seg 1/1 to: AF153283 from: 1 to: 930  
 28 SerProLeuAsnLeuLeuLeuLeuGlyLeuCysIlePheLeuLeuTyrlLy 44



```
||||| : : : : : ||| : : : : :
69 TCTCCAGCTGCGTTTTTTCACCGTACTTGTCTCGCTTTCGCCGTTTACCA 118
44 stleValArgGly.....AspGlnProAlaAlas 54
119 ACTCGTCTCCGTTCTTCTCGTTCTCTCCTGAAAGTTTCCACCGACCTCGGTCTT 168
54 exGlyAspArgThrThr***ProProProLeuProArgLeuLysArg 70
169 TGGAGTTTCAGCTCAATCGGAGCTCTTCCA...CCGCCGTTTCAGCTC 215
71 ArgAspPheThrProAlaGluLeuArgArgPheAspGlyValGlnAspPr 87
216 GGAGAAATCACTGAGGAGGAGCTTAAGCTTTATGATGGCTCCGATCTAA 265
87 ohrg...IleLeuMetAlaIleAsnGlyLysValPheAspValThrLysG 103
266 AAAGCCCTTCTTATGGCGATCAAGGCGCAGATCTATGATGTTTCTCAGA 315
103 LVArgLysPheTyrGlyProGluGlyProTyrGlyValPheAlaGlyArg 119
316 GCAGGATGTTTATGACCAAGTGGCGCATATCTCTGTTTGAGGGA 365
120 AspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluAlaLeuLy 136
366 GATGCAAGCGAGCTCTGCAAGATGTCTATTGAGACCAA..... 407
136 shaspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrL 153
408 .GACTTGACTGGAGACATCTCAGGTCTCGGTGCTATTGAGTAGAGCGT 456
153 euSerAspPrpGluSerGlnPheThrPheLysTyrHisHisValGlyLys 169
457 TACAAGCTGGAGTACAAAGTTCATGAGCAAGTATGTCAAGTCGGAACC 506
170 LeuLeuLysGluGlyGluGluProThrValTyrSerAspGluGluPr 186
507 ATTCAAAGAAGAGTGAGCAAGCAAGAAAGTTTCAAGCTCCGNAAGC 556
186 oLysAspGluSerSer 191
557 AAAGACTGCTCTGCG 572
```

seq\_name: gb\_p12:AF153284

seq\_documentation\_block:
LOCUS AF153284 847 bp mRNA PLN 02-JUN-1999
DEFINITION Arabidopsis thaliana putative progesterone-binding protein homolog
(ATmpl) mRNA, complete cds.

ACCESSION AF153284
VERSION AF153284.1 GI:4960155

KEYWORDS
SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

REFERENCE
1 (bases 1 to 847)
Choi J.H., Choi H. and Gray, P.

AUTHORS Plant homologues of mammalian putative progesterone-binding
membrane proteins

JOURNAL Unpublished
REFERENCE
2 (bases 1 to 847)

AUTHORS Choi J.H., Choi H. and Gray, P.
TITLE Direct Submission

JOURNAL Submitted (19-MAY-1999) School of Biology, Georgia Institute of
Technology, MC 0230, Atlanta, GA 30332-0230, USA

FEATURES
Location/Qualifiers
1..847

source
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db\_xref="taxon:3702"

```
/db_xref="dbEST:N37902"
/Chromosome="5"
1..847
/gene="Atmpl"
62..724
/gene="Atmpl"
/codon_start=1
/product="putative progesterone-binding protein homolog"
/protein_id="AAD34616.1"
/db_xref="GI:4960156"
/translation="MALELWQLKEAIHAYTGLSPVFETLALAFATYQVIVSGWFPAS
PFDDVNRHQRARSIAQEEPEPPQVQVEITEELKQYDGDSPQKPLLMALIKHIIYD
VTQSRMFYGGPGPYALFAGKASRALAKMSFEKDLTWDISGLGPFELDALQWYKFE
MSKYAKVCTVKVAGSEPTASVSEPTENVEQDAHVTTTPEKTVVDKDDAPAEIVLKK
EE"
BASE COUNT 237 a 185 c 192 g 233 t
ORIGIN
alignment_scores:
Quality: 241.00 Length: 177
Ratio: 2.231 Gaps: 6
Percent Similarity: 61.017 Percent Identity: 34.463
alignment_block:
US-09-203-548-1 x AF153284 ..
Align seg 1/1 to: AF153284 from: 1 to: 847
```

```
20 GlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuLeuGlu 36
113 GGCTTTCTCTGTTGCTTCTTCTCAGCTGCTAGCTCTC..... 151
36 yLeuCysIlePheLeuLeuTyrLysIleValArgGly..... 48
152 .....GCCTTCGCCATTTACCAAGTCATCTCAGGCTGTTGCCCTCGC 194
49 .....AspGlnProAlaSerGlyAspArgThrThrThr 60
195 CGTTCGATGATGTTAACCGACATCAGAGCATGATCCTTGGCTCAAGAG 244
61 ***ProProProLeuProArg...LeuLysArgArgAspPheThrProAl 76
245 GAGGAGCCACCGATCTCAGCTGTTCAAGTCGGTGAGATCAGCGAGGA 294
76 aGluLeuArgArgPheAspGlyValGlnAspProArg.....IleLeuM 91
295 GGAGCTTAACAGTACATGATGCG...TCTGATCTCAAAAGCCCTTCTTA 341
91 etAlaIleAsnGlyLysValPheAspValThrLysGlyArgLysPheTyr 107
342 TGGCTATCAACATCAGATCTATGATGTTACACAAGCAGGATGTTCTAC 391
108 GlyProGluGlyProTyrGlyValPheAlaGlyArgAspAlaSerArgG 124
392 GGACGAGGAGGCGCATATGTTGTTTGCAGGGAAGACGCTAGCCGAGC 441
124 yLeuAlaThrPheCysLeuAspLysGluAlaLeuLysAspGluTyrAsp 141
442 TCTTGCAAAAGATGTCATTTGAGGAG.....AAAGACTTGACTTGGG 482
141 spLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeuSerAspTrpGlu 157
483 ATATCTCTGGTCTTGGTCCCTTTGAGCTAGATGCTCTTCAAGATTGGAG 532
158 SerGlnPheThrPheLysTyrHisHisValGlyLysLeuLeuLysGlu 174
533 TACAAGTTCATGAGCAAGTATGTAAGTTGGTACTGTCTCAAGTGGCTGG 582
174 yGluGluProThrValTyrSerAspGluGlu 184
583 TTCAGAACCTTGAAACCCGCATCTGTCTCTGAA 613
```



OM of: US-09-203-548-1 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Apr 19, 2000 3:54 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool/US09203548/runat\_14042000\_104705\_23262/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=ring -GAPOF=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOF=4.500 -OGAPEXT=0.050 -XGAPOF=10.000 -XGAPEXT=0.500  
-FGAPOF=6.000 -FGAPEXT=7.000 -YGAPOF=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcpt  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=100000 -USER=US09203548 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

Search information block:

Query: US-09-203-548-1

Query length: 220

Database: N\_Geneseq\_36:\*

Database sequences: 311585

Database length: 125096042

Search time (sec): 32.650000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:V60245	+	1121.00	2355.79	13e-123	788
N_Geneseq_36:V84368	+	1068.00	2233.94	7.9e-117	1875
N_Geneseq_36:V63170	+	1061.00	2219.13	5.3e-116	1868
N_Geneseq_36:V09852	+	999.50	2088.53	9.9e-109	1893
N_Geneseq_36:V09878	+	501.00	1026.99	1.3e-49	2776
N_Geneseq_36:V69409	+	88.50	170.05	0.0717	530
N_Geneseq_36:V61885	+	85.50	136.98	4.98	6103
N_Geneseq_36:V62025	+	84.50	148.76	1.10	1719
N_Geneseq_36:V68206	+	80.50	153.52	0.5973	509
N_Geneseq_36:Q41259	+	80.50	135.61	5.94	2639
N_Geneseq_36:V72718	+	80.50	135.61	5.94	2639
N_Geneseq_36:V86375	+	79.50	102.91	393.39	43804
N_Geneseq_36:V6690	+	79.50	102.91	393.39	43804
N_Geneseq_36:V60472	+	79.50	126.40	19.34	4590
N_Geneseq_36:V32232	+	77.50	114.60	87.93	10140
N_Geneseq_36:V07170	+	77.00	129.47	13.05	2345
N_Geneseq_36:V0707	+	77.00	129.06	13.76	2435
N_Geneseq_36:V83172	+	76.00	122.26	32.89	3742
N_Geneseq_36:V69503	+	75.00	138.83	3.93	672
N_Geneseq_36:Q25811	+	75.00	111.37	132.99	8378
N_Geneseq_36:Q46542	+	74.50	129.44	13.10	1444
N_Geneseq_36:V09623	+	74.50	129.44	13.10	1444
N_Geneseq_36:Q46541	+	74.50	126.47	19.18	1898
N_Geneseq_36:V09622	+	74.50	126.47	19.18	1898
N_Geneseq_36:Q20576	+	74.50	123.34	28.63	2529
N_Geneseq_36:V63209	+	74.50	123.15	29.34	2574
N_Geneseq_36:V09621	+	74.50	122.09	33.61	2837
N_Geneseq_36:Q46540	+	74.50	122.09	33.65	2839
N_Geneseq_36:Q86294	+	74.00	115.52	78.07	4707
N_Geneseq_36:V32273	+	73.50	126.04	20.27	1625
N_Geneseq_36:V69761	+	73.50	125.96	20.48	1637
N_Geneseq_36:Q3018	+	73.50	124.96	23.27	1794
N_Geneseq_36:Q35926	+	73.50	122.95	30.10	2157
N_Geneseq_36:V2935	+	73.50	122.95	30.10	2157
N_Geneseq_36:Q2820	+	73.50	122.95	30.12	2158
N_Geneseq_36:T84865	+	73.00	124.35	25.18	1722
N_Geneseq_36:V328415	+	73.00	123.63	27.60	1839
N_Geneseq_36:V28424	+	73.00	123.31	28.76	1894
N_Geneseq_36:V9337	+	72.50	123.92	26.61	1625
N_Geneseq_36:V32790	+	72.50	117.13	63.55	3032
N_Geneseq_36:V32173	+	72.50	100.27	552.24	14273

N\_Geneseq\_36:T13486 + 72.00 125.52 21.66 1272 ! Rice NADH-dependent reducta  
N\_Geneseq\_36:Q34549 - 72.00 121.10 38.20 1910 ! Streptomyces lividans prote  
N\_Geneseq\_36:Q43965 + 72.00 114.76 86.13 3420 ! Pyrococcus thernostable DNA  
N\_Geneseq\_36:X20506 + 71.50 131.83 9.64 646 ! Polynucleotide sequence from  
seq\_name: N\_Geneseq\_36:V60245

seq\_documentation\_block:

ID V60245 standard; DNA; 788 BP.

AC V60245;

DT 14-DEC-1998 (first entry)

DE Nucleotide sequence encoding human cytokine/steroid receptor protein.

ss: human; cytokine/steroid receptor protein; CYSTAR;

KW developmental disorder; aberrant cellular differentiation; inflammation.

OS Homo sapiens.

FT Key

FT CDS

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

alignment\_scores:

Quality: 1121.00 Length: 220

Ratio: 5.190 Gaps: 0

Percent Similarity: 98.182 Percent Identity: 98.182

alignment\_block:

US-09-203-548-1 x V60245

Align seg 1/1 to: V60245 from: 1 to: 788

1 MetalaalaGluAspValValAlaThrGlyAlaAspProSerAspLeuCl 17

100 ATGGCTCCGAGGATGTGTGGCGACTGGCGCCGACCCAGCGATCTGGA 149

17 userGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeu 34

150 GAGCGGGGGGGTGTGTGATGATGATTTTCACGCGCGCTCAACCTGCTGC 199

34 euLeuGlyLeuCysIlePheLeuLeuLeuLeuLeuLeuLeuLeuLeu 50

200 TGTGTGGCTCTGCATCTCTCTGCTCTACAAGATCGTCGCGGGGACGAC 249

51 ProLaalaSerGlyAspArgThrThrThr\*\*\*ProProProLeuProAr 67

250 CCGGCGGCGCAGCGGACGACGACGACGACGACGACGACGACGACGACG 299

```

67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
|||||
300 CCTCAAGCGCGCGACTTCACCCCGCCGAGCTGGCGCCTTCGACGGCG 349
|||||
84 alGlnAspProArgLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
350 TCCAGGACCGCGCATCTCATGGCCATCAACGGCAAGGTGTTGCGATGTG 399
|||||
101 ThrLysGlyArgLysPheThrGlyProGluGlyProTyrGlyValPheAl 117
|||||
400 ACCAAAGCCGCAAAATTCACGGCCGCGAGGCGCGTATGGGTCTTTGC 449
|||||
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
450 TGGAAAGATGATGCATCCAGGGCGCTTCCACATTTTGCCTGGATAAGAA 499
|||||
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
500 CACTGAAGGATGAGTACGATGACCTTTCGACCTCACTGCTGCCACGAG 549
|||||
151 GluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHisVa 167
|||||
550 GAGACTCTGAGTGACTGGGAGTCTCAGTTCACTTTCAAGTATCATCACGT 599
|||||
167 lGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspGlu 184
|||||
600 GGGCAAACTGCTGAAGAGGGGAGGAGGCCACTGTGTACTCAGATGAGG 649
|||||
184 luGluProLysAspGluSerArgLysAsnValLysAlaPheSerGly 200
|||||
650 AAGAACCAAAAGATGAGAGTTCCCGGAAAAATGTTAAACCAATTCAGTG 699
|||||
201 SerIleSer*****TyrPheAlaLysSerPheValThrValHis***Va 217
|||||
700 AGTATATCATNTNTGATTTTCAAAATCATTTGTAACAGTCCACATNTGT 749
|||||
217 lPheLysThr 220
|||||
750 CTTTAAACA 759

```

seq\_name: N\_Geneseq\_36:V84368

seq\_documentation\_block:

ID V84368 standard; cDNA to mRNA; 1875 BP.

AC V84368;

DT 30-MAR-1999 (first entry)

DE Human stomach carcinoma cDNA clone HP10413.

KW Transmembrane protein; HP10413; human; stomach cancer; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 79..666

FT /\*tag= a

FT /note= "cDNA comprising the coding region (minus the stop codon) is claimed (claim 3)"

PN WO9855508-A2.

PD 10-DEC-1998.

PF 03-JUN-1998; J02445.

PR 03-JUN-1997; JP-144948.

PA (PROT-) PROTEGENE INC.

PA (SAGA ) SAGAMI CHEM RES CENTRE.

PI Kato S, Sekine S, Yamaguchi T.

DR WPI; 99-045730/04.

DR P-PSDB; W88500.

PT New human proteins containing transmembrane domains and their

PT encoding sequences - useful in the preparation of antibodies and

PT large-scale protein production, gene diagnosis, and gene therapy

PS Claim 4; Page 138-140; 178pp; English.

CC This is the nucleotide sequence of cDNA clone HP10413, which

CC includes a coding region (also claimed) for a novel human

CC transmembrane protein (see W88500). The clone was isolated from a

CC stomach cancer cDNA library using a signal sequence detection

CC method, and by protein synthesis by in vitro translation. The

CC encoded protein (28 kDa) has one transmembrane domain at the  
 CC N-terminal. It shows 96.4% homology to the pig steroid  
 CC membrane-binding protein. The invention provides nucleotide  
 CC sequences (see W84359-76) coding for 18 transmembrane proteins  
 CC (see W88491-508), vectors containing such polynucleotides, and  
 CC eukaryotic cells containing the vectors. The proteins can be  
 CC used as antigens or as compositions in the preparation of  
 CC antibodies against the proteins. The polynucleotides can be used  
 CC as probes for gene diagnosis, and as gene sources for gene therapy  
 CC and large-scale production of proteins encoded by the cDNA. The  
 CC host cells are used for the detection of ligands corresponding to  
 CC the expressed proteins, and the screening of low mol.wt. medicines.  
 CC Sequence 1875 BP; 532 A; 396 C; 426 G; 521 T;

alignment\_scores:

Quality: 1068.00 Length: 222

Ratio: 5.014 Gaps: 2

Percent Similarity: 95.946 Percent Identity: 95.045

alignment\_block:

US-09-203-548-1 x V84368 ..

Align seg 1/1 to: V84368 from: 1 to: 1875

```

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG 17
79 ATGGCTGCGGAGGATGTGTGGCGACTGGCGCGACCCCAAGCGATCTGGA 128
17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeu 34
129 GAGCGCGGGCTGCTGATGAGATTTTACGTGCGCGCTCAACCTGCTGC 178
34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50
179 TGCTTGGCTCTGCATCTTCTCTCTACAAGATCGTGGCGGGGACCAG 228
51 ProAlaAlaSerGlyAspArgThrThr***ProProProLeuProAr 67
229 CCGGGCGCGGCGGCGGACGACGACGACGACGCGCGCGCGCTTGCCTG 278
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
279 CCTCAAGCGCGCGGACTTCACCCCGCGAGCTGGCGCGCTTCGACGGCG 328
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
329 TCCAGGACCGCGCATACTCATGGCCATCAACGGCAAGGTGTTGCGATGT 378
101 ThrLysGlyArgLysPheThrGlyProGluGlyProTyrGlyValPheAl 117
379 ACCAAAGCCGCAAAATTCACGGCCGCGAGGCGCGTATGGGTCTTTGC 428
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGlu 134
429 TGGAAAGATGATGCATCCAGGGCGCTTCCACATTTTGCCTGGATAAGAA 478
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
479 CACTGAAGGATGAGTACGATGACCTTTCGACCTCACTGCTGCCACGAG 528
151 GluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHisVa 167
529 GAGACTCTGAGTGACTGGGAGTCTCAGTTCACTTTCAAGTATCATCACGT 578
167 lGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspGlu 184
579 GGGCAAACTGCTGAAGAGGGGAGGAGGCCACTGTGTACTCAGATGAGG 628
184 luGluProLysAspGluSerArgLysAsnValLysAlaPheSerG 200
629 AAGAACCAAAAGATGAGAGTGGCGGAAAAATGATTAAAGCATTCAGTGG 678

```



CC diagnosis and therapy of tumour cells in organs targeted by  
CC steroid hormones and/or opioids, and nerve cells. Probes and  
CC primers based on the receptor cDNA can be used to detect expression  
CC of the receptor, and isolate related coding sequences. It can  
CC also be used to produce recombinant polypeptides.  
SQ Sequence 1893 BP; 509 A; 447 C; 461 G; 476 T;

alignment\_scores:  
Quality: 99.50 Length: 222  
Ratio: 4.782 Gaps: 3  
Percent Similarity: 94.144 Percent Identity: 89.640

alignment\_block:

US-09-203-548-1 x V09852 ..

Align seg 1/1 to: V09852 from: 1 to: 1893

```
1 MetAlaLaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17
|||||
42 ATGGCTGCGGAGGATGTGGCGGTACCGCGCGGACCGCGAGCGAGCTAGA 91
|||||
17 userGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34
|:::|||||
92 GGGCGGGGGCTGCTGCATGAGATTTTCACGTCGCGGCTCAACCTGCTGC 141
|||||
34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50
|||||
142 TGCTCGGCTCTGCATCTTCCTGCTCTACAGATCGTGGCGGGGACCA 191
|||||
51 ProAlaLaSerGlyAspArgThrThr***ProProProLeuProAr 67
|||||
192 CCGCGCGCCAGC...GATACGACGACGACGAGCGCCCGCGTGGCCCG 238
|||||
67 gluLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
|||||
239 CCTTAAGCGCGCGACTTCACCGCTGCCGAGCTCGCTCGCTTCGACGGCG 288
|||||
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
289 TCCAGAGCCCGGTATACATCATGGCCATCAACGGCAAGGTGTTCGACGTG 338
|||||
101 ThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPheAl 117
|||||
339 ACCAAGCGCCGAAGTTCTACGGCCCGGAGCGCGTACGGGGTCTTTGC 388
|||||
117 acGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGlu 134
|||||
389 TGGAGAGACGCATCCAGGGCGCTGGCCACGTTTTCCTGGATAAGGAAG 438
|||||
134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaGlnGln 150
|||||
439 CCCTGAAGACGAGATGATGACCTTCTGACCTCACTCCTGCCACGAC 488
|||||
151 GluThrLeuSerAspTyrGluSerGlnPheThrPheLysTyrHisHisVa 167
|||||
489 GAGACCTGATGATGATGGGACTCTCAGTTCACTTCAAGTACATCACGT 538
|||||
167 lGlyLysLeuLeuLysGluGlyGluProThrValTyrSerAspGluG 184
|||||
539 GGGCAAACTGCTGAAGGAGGGGAGGAGCCACCGTGTACTCAGATGAGG 588
|||||
184 luGluProLysAspGluSerArgLysAsn.ValLysAlaPheSerG1 200
|||||
589 AAGAGCCCAAGATGAGAGCGCTCGGAAATGATTAAACCGCTTCGGTG 638
|||||
200 ySerIleSer.*****TyrPheAlaLysSerPheValThrValHis*** 216
|||||
639 AAGCATATCTATTTTGTGAGAAATCATTTGTAAACATTCAGTCT 688
|||||
217 ValPheLysThr 220
|||||
689 GTCCTAAACA 700
```

seq\_name: N\_Geneseq\_36:X00678

seq\_documentation\_block:

ID X00678 standard; DNA; 2776 BP.

AC X00678;  
DT 25-MAR-1999 (first entry)  
DE Human secreted protein gene 68 clone HLB049.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; ischaemia; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN WO9842738-Al.  
PD 01-OCT-1998.  
PF 19-MAR-1998; U05311.  
PR 30-MAY-1997; US-050937.  
PR 21-MAR-1997; US-041276.  
PR 21-MAR-1997; US-041277.  
PR 21-MAR-1997; US-041281.  
PR 21-MAR-1997; US-042344.  
PR 30-MAY-1997; US-048069.  
PR 30-MAY-1997; US-048094.  
PR 30-MAY-1997; US-048095.  
PR 30-MAY-1997; US-048096.  
PR 30-MAY-1997; US-048099.  
PR 30-MAY-1997; US-048131.  
PR 30-MAY-1997; US-048135.  
PR 30-MAY-1997; US-048154.  
PR 30-MAY-1997; US-048160.  
PR 30-MAY-1997; US-048186.  
PR 30-MAY-1997; US-048187.  
PR 30-MAY-1997; US-048188.  
PR 30-MAY-1997; US-048350.  
PR 30-MAY-1997; US-048351.  
PR 30-MAY-1997; US-048352.  
PR 30-MAY-1997; US-048355.  
PR 05-AUG-1997; US-054804.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA,  
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,  
PI Rosen CA, Ruben SM, Shi Y, Young P;  
DR WPI; 99-070066/06.  
P-PSDB: W67874.  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 1; Page 227-228; 385pp; English.  
CC This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. X00602) for increasing the stability of the fused protein as  
CC compared to the human protein only.  
CC The invention relates to 87 novel genes and their fragments (nucleic acid  
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 87 polynucleotides, based on  
CC which tissues they are most highly expressed in (see X00611 for described  
CC uses).  
SQ Sequence 2776 BP; 836 A; 514 C; 623 G; 799 T;

alignment\_scores:

Quality: 501.00  
Ratio: 3.767

Length: 167  
Gaps: 1











120 pAlaSerArgGlyLeuAlaThrPhe..... 128  
 1580 CGCATCGAGATCCTCGCGAGCGCTTTCGGCGAACTCCGCATAAGCCT 1629  
 129 .....CysLeuAspLysGluAlaLeuLysAspGluTyrAsp 140  
 1630 CGTGTTCGAAGCCATGCATGTGGCGACGCGTGTGTCGGCGAAA..... 1673  
 141 AspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeuSerAspTrpG1 157  
 1674 CGCGTGTCCGGGTGCACGGCGCGCACACGAGATGCGCGCACGCCGA 1723  
 157 userGlnPheThrPheLysTyrHisHisValGlyLysLeuLysGluG 174  
 1724 A...CGGTTCGATTCGACCGCCATCATTCGGTGAACGATTCAGCCGCC 1770  
 174 LyGluGluProThrValTyrSerAspGluGluGluProLysAspGluSer 190  
 1771 GCCTTGTGCGCGGTATGCGCGGACCGACGCGGCAACACCTTCACCGTGAC 1820  
 191 SerArgLysAsnValLysAlaPheSerGly 200  
 1821 GCTCGACGTGACGTTACGATCAGCGCGGC 1850

seq\_name: N\_Geneseq\_36:T86375

seq\_documentation\_block:

ID T86375 standard; DNA; 43804 BP.  
 AC T86375;  
 DT 01-MAY-1998 (first entry)  
 DE Chicken embryo lethal orphan virus genome.  
 KW Genetic modification; gene therapy; cancer vaccine; vaccine;  
 KW Infectious disease; ss.  
 OS Chicken embryo lethal orphan virus.  
 PN DE19615803-A1.  
 PD 23-OCT-1997.  
 PF 20-APR-1996; 015803.  
 PR 20-APR-1996; DE-015803.  
 PA (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
 PI Baker A, Chiocca S, Cotten M, Kurzbaue R, Schaffner G;  
 DR WPI; 97-514518/48.  
 PT Genetically modified CELO viruses - useful for gene therapy or  
 vaccine production, e.g. against cancer  
 PS Claim 1; Fig 5; Sipp; German.  
 CC The present sequence is the chicken embryo lethal orphan (CELO)  
 CC virus genome, which was used in the preparation of CELO virus  
 CC obtainable by manipulation of CELO virus DNA in vitro. Modified  
 CC CELO viruses containing exogenous DNA encoding a therapeutic  
 CC protein are useful for gene therapy. Modified CELO viruses  
 CC containing exogenous DNA encoding an immunostimulant protein  
 CC (especially a cytokine) or a tumour-associated antigen can be used  
 CC to produce cancer vaccines. Modified CELO viruses containing  
 CC exogenous DNA encoding an antigen derived from a human, animal or  
 CC avian pathogen can be used to produce vaccines against infectious  
 CC diseases of humans, animals and birds, respectively.  
 SQ Sequence 43804 BP; 10285 A; 12086 C; 11698 G; 9735 T;

alignment\_scores:

Quality:	79.50	Length:	202
Ratio:	0.914	Gaps:	11
Percent Similarity:	43.069	Percent Identity:	26.238

alignment\_block:

US-09-203-548-1 x T86375 ..

Align seg 1/1 to: T86375 from: 1 to: 43804

51 ProAlaAlaSerGlyAspArgThrThr\*\*\*ProProLeuProAr 67

|||||:|||||  
 20621 CCNACTTCCTGGACACCGCTGTACGACCTGGTGTCCAGCAGCTCCCGTT 20670

67 gLeuLysArgArgAspPheThrProAlaGluLeuArg...ArgPheAspG 83  
 |||||:|||||:|||||  
 20671 GTTAAACG.....ACACCGGCTCACAGCGCGTCTCAGGACAACGT 20708  
 83 lyValGlnAspProArgIle.....LeuMetAlaIleAsnGlyLysVal 97  
 |||||:|||||:|||||  
 20709 CGCTAAACAACCTCCGGCTTTATCGCCCTCGCAGCT..... 20743  
 98 PheAspValThrLysGlyArgLysPheTyrGlyProGluGlyProTyrG1 114  
 20744 .....GCCCCGTATGG 20754  
 114 yValPhe AlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLe 130  
 |||||:|||||:|||||:|||||:|||||:|||||  
 20755 ACCCGACAGCAGGCGGAGCGCTGCGCGCTAACTGGCGGTACCGCTGAT 20804  
 130 uAspLysGluAlaLeuLysAspGluTyrAspAspLeuSerAspLeuThrA 147  
 :|||||:|||||:|||||:|||||:|||||:|||||  
 20805 CGGGAACGACGCCATCAGT.....TCCAACCAAAACCG 20836  
 147 laAlaGlnGlnGluThrLeuSerAsp.....TrpGluSerGlnPhe 160  
 :|||||:|||||:|||||:|||||:|||||  
 20837 TCAACTACAAGAGTTCCTGTGCGATAACTACCTCTGGACGCTGCCGTTC 20886  
 161 ThrPheLysTyrHisHisValGlyLysLeuLysGluGlyGluGluPr 177  
 :|||||:|||||:|||||:|||||:|||||:|||||  
 20887 AGCTCGGACTTTATGTATATGGGAGAGCTGACCGATCTGGGTGAGAACCC 20936  
 177 o.....ThrValTyrSerAspGluGluGluPr 186  
 |||||:|||||:|||||:|||||:|||||  
 20937 CATGTACACAACAACCTCCCATAGCATGGTTATCAACTTTCAGTTGGACC 20986  
 186 roLysAspGluSerSer..... 191  
 |||||:|||||:|||||:|||||:|||||  
 20987 CCATGGATGAGATACTACCTGTACATGCTGTACGGGGTATTTGATACC 21036  
 192 .....ArgLysAsnValLysAlaPheSerGlySerIl 202  
 :|||||:|||||:|||||:|||||:|||||  
 21037 GTTCGCGTGAACCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21078  
 202 eSer\*\*\*\*\*TyrPheAlaLysSerPheValThrValHis\*\*\*ValPheL 219  
 |||||:|||||:|||||:|||||:|||||  
 21079 .....TACTTCGCTAGCCCTTTCGCCACAGGCAACCGCTGTGTATAA 21118  
 219 ys 219  
 ||  
 21119 AA 21120

seq\_name: N\_Geneseq\_36:X26690

seq\_documentation\_block:

ID X26690 standard; DNA; 43804 BP.  
 AC X26690;  
 DT 18-JUN-1999 (first entry)  
 DE Complete genome sequence of avian adenovirus CELO.  
 KW Chicken embryo lethal orphan; CELO; adenoviral vector; CELO adenovirus;  
 KW vaccine; Marek's disease; infectious bronchitis; laryngotracheitis;  
 KW gumbo disease; Newcastle disease; ds.  
 OS Avian adenovirus.  
 PN WO9909194-A1.  
 PD 25-FEB-1999.  
 PF 13-AUG-1998; F01803.  
 PR 14-AUG-1997; FR-010386.  
 PA (CNEV-) CNEVA CENT NAT ETUD VETERINAIRES & ALIM.  
 PA (NAVE-) CENT NAT ETUD VETERINAIRES & ALIMENTAIRE.  
 PI Langlois P;  
 DR WPI; 99-181049/15.  
 PT Production of recombinant CELO adenovirus vectors - by inserting  
 PT heterologous DNA into nonessential region of CELO genome  
 PS Claim 4; Fig 1; 121pp; French.  
 CC The present sequence represents the complete genome sequence of avian  
 CC adenovirus chicken embryo lethal orphan (CELO). The specification  
 CC describes a method for preparing recombinant CELO adenovirus vectors,

CC which comprises inserting a DNA sequence encoding a heterologous  
 CC polypeptide into a nonessential region of the CELO adenovirus genome.  
 CC The adenoviral vectors can be used to produce vaccines. The vaccine  
 CC can be used to protect birds against infectious diseases, especially  
 CC Marek's disease, infectious bronchitis, laryngotracheitis, gumbo  
 CC or Newcastle disease. The vector can be used to produce a recombinant  
 CC polypeptide of interest in vivo by infecting an animal with the vector,  
 CC in which the inserted DNA sequence encodes the polypeptide of interest,  
 CC and recovering the polypeptide from the animal.  
 SQ Sequence 43804 BP; 10284 A; 12086 C; 11700 G; 9734 T;

alignment\_scores:  
 Quality: 79.50 Length: 202  
 Ratio: 0.914 Gaps: 11  
 Percent Similarity: 43.069 Percent Identity: 26.238

alignment\_block:

US-09-203-548-1 x X26690 ..

Align seg 1/1 to: X26690 from: 1 to: 43804

51 ProAlaIaSerGlyAspArgThrThr\*\*\*ProProLeuProAr 67

20621 CCAACTTCTCGACACACCGCTGTACGACCTGGTGTCCAGCACTCCGTT 20670

67 gLeuLysArgArgAspPheThrProAlaGluLeuArg....ArgPheAspG 83

20671 GTTACG.....ACACCGGCTCACAGCGCTCTCAGCAACAACGT 20708

83 lyValGlnAspProArgIle.....LeuMetAlaIleAsnGlyLysVal 97

20709 GGGTAACAACCTCGGCTTTATCGCCCTCGCAGCT..... 20743

98 PheAspValThrLysGlyArgLysPheThrGlyProGluGlyProTyrGln 114

20744 .....GGCCGCTATGG 20754

114 yValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCys.Le 130

20755 ACCGCACAGCAGGGGAGCCCTGGCCCGTAACCTGCGCGTACCCGCTGAT 20804

130 uAspLysGluAlaLeuLysAspGlyTyrAspAspLeuSerAspLeuThrA 147

20805 CGGGAACGCGCCATCAGT.....TCCAACCAACCG 20836

147 laAlaGlnGlnGluThrLeuSerAsp.....TrpGluSerGlnPhe 160

20837 TCAACTACAAGAAGTCTCTGTGCGATACTACCTCTGACCGCTGCCGTTTC 20886

161 ThrPheLysTyrHisHisValGlyLysLeuLysGluGlyGluGluPhe 177

20887 AGCTCGGACTTTATGTATATGGAGACGCGCGATCTCGGTCAGAACCC 20936

177 o.....ThrValTyrSerAspGluGluGluPhe 186

20937 CATGTACACAACAACACTCCCATAGCATGTTTATCACTTTGATGGACC 20986

186 roLysAspGluSerSer..... 191

20987 CCATGGATGAGAATACTACGTGTACGTGTACGCTGTACGGGTATTTGATACC 21036

192 .....ArgLysAsnValLysAlaPheSerGlySerIle 202

21037 GTTCCCGTGAACACCGCCGAGGCTAACGTGCTAGCCATGGCT..... 21078

202 eSer\*\*\*\*TyrPheAlaLysSerPheValThrValHis\*\*\*ValPheI 219

21079 .....TACTTCGGTACCGCTTTCCCCACAGCAACGCTGTGTAAA 21118

219 ys 219

21119 AA 21120

seq\_name: N\_Geneseq\_36:N60472

seq\_documentation\_block:

ID N60472 standard; DNA; 4590 BP.

AC N60472;

DE 24-AUG-1991 (first entry)

DE Sequence encoding the ring-infected Erythrocyte Surface Antigen

DE (RESA).

KW Malaria vaccine; antigen; epitope; ss.

OS Plasmodium falciparum.

FH key Location/Qualifiers

FT exon 801..995

FT exon 1199..4225

FT exon /\*tag= a

FT exon /\*tag= b

PN W08601802-A.

PD 27-MAR-1986.

PF 11-SEP-1985; 006960.

PR 11-SEP-1984; AU-007067.

PR 11-SEP-1984; AU-007066.

PR 10-SEP-1985; AU-047326.

PA (HALL-) HALL INST MED RES.

PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;

DR WPI; 86-094065/14.

DR P-PSDB; P60569.

PT DNA coding for Plasmodium falciparum antigens - expressing

PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P

PT falciparum

PS Claim 4; Fig 1; 55pp; English.

CC The inventors claim a novel DNA molecule which comprises a

CC nucleotide sequence corresp. to all or a portion of the base

CC sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have

CC antigenicity suitable for providing protective immunity against

CC Plasmodium falciparum malarial infections.

SQ Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;

alignment\_scores:

Quality: 79.00 Length: 131

Ratio: 1.215 Gaps: 7

Percent Similarity: 49.618 Percent Identity: 25.954

alignment\_block:

US-09-203-548-1 x N60472 ..

Align seg 1/1 to: N60472 from: 1 to: 4590

65 LeuProArgLeuLysArgArgAspPheThrProAlaGluLeuArgArgPh 81

2129 TTACCATCATTA...AGGCCAGTATTACTAATTCAGCTATTAAATTATTA 2175

81 eAspGlyValGlnAspProArgIleLeuMetAlaIleAsnGlyLysValP 98

2176 TGATACCGTAAAGAT..... 2191

98 heAspValThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGly 114

2192 .....GGT 2194

115 ValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAs 131

2195 GTATACCTTAGACCATGAACATCAGATGCTCTTTATACA.....GA 2235

131 pLysGluAlaLeu.....LysAspGluTyrAspLeuSerA 144

2236 TGAAGATTGTTTATTGATTTGAAAAACAAAATATATGATATGTTAG 2285

144 sPleuThrAlaAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPhe 160

2286 AT.....ACATCTGAAGAAGAATCTGTTGAAGAAAAATGAAGAAGACAC 2329

161 ThrPheLysTyrHisHisValGlyLysLeuLysGluGly..... 174

```
||||
2330 ACTGTTGATGATGAACATGTAGAGAACACACACTGCTGATCAGCAACATGT 2379
175 .GluGluProThrValTyrSerAspGlu.....GluGluPro 186
|||||
2380 AGAAGAACAACCATGTTGCTGATGATGAACATGTAGAGAACA 2422
seq_name: N_Geneseq_36:T32232
seq_documentation_block:
ID T32232 standard; DNA; 10140 BP.
AC T32232;
DT 27-OCT-1996 (first entry)
DE Plasmid pBE92
KW Thermotable enzyme; xylanase; xynA gene; beta-glucosidase;
KW Bacillus; Caldocellum saccharolyticum; polymerase chain reaction;
KW PCR; primer; signal peptide; pBE92; ss.
OS Chimeric Bacillus sp.;
PN Chimeric synthetic.
PN WO9623887-A1.
PD 08-AUG-1996.
PF 24-JAN-1996; U00891.
PR 30-JAN-1995; US-380521.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI Jackson EN, Liu G;
DR WPI: 96-371435/37.
PT Thermostable enzymes e.g. xylanase and beta-glucosidase - are
PT produced extracellularly, in absence of generally required signal
PT peptide, and in high yields from recombinant bacteria
PS Example 1: Page 32-38; 48pp; English.
CC Plasmid pBE92 (T32232) contains the Bacillus alkaline protease
CC gene promoter (aprp) and signal sequence (aprss). A PCR
CC product (see also T32230-31) contg. the Caldocellum saccharolyticum
CC xylanase gene (xynA) was ligated to the large fragment of pBE92.
CC The resulting plasmid, pBEI58, contained the aprp-aprss-xynA fusion.
CC Bacillus subtilis transformants produced xylanase at 37 U/ml
CC supernatant; this compared with 272 and 1110 U/ml using constructs
CC aprp-xynA (pBEI45) (see also T32225-26) and nprp-xynA (pBEI45) (see
CC also T32227), respectively, which lacked a signal sequence.
SQ Sequence 10140 BP; 2821 A; 2414 C; 2129 G; 2776 T;
```

```
alignment_scores:
Quality: 77.50 Length: 171
Ratio: 0.912 Gaps: 9
Percent Similarity: 49.708 Percent Identity: 24.561
alignment_block:
US-09-203-548-1 x T32232 ..
Align seg 1/1 to: T32232 from: 1 to: 10140
33 LeuLeuGlyLeuGlyLeuPheLeuLeuTyrLysIleValArgGlyAs 49
|||||
1142 TTGCTGTTGCTTTAGCG.....TTAATCTTTACGATGCGGCTTCGGCAG 1185
49 pGlnProAlaAlaSerGlyAspArgThrThr***ProProProLeup 66
:|||||
1186 CACATCTCTGCTAGCGCGATATCCGGACA.....CCAGAAATGC 1226
66 roArgLeuLysArgArg.....AspPheThr...ProAlaGlu 77
|||
1227 CTGTTCTGGAACACCGGCTGCTCAGGCGGATATTACTGCACCGCGCGGT 1276
78 LeuArgArgPheAspGlyValGlnAspProArgIleLeuMetAlaIleAs 94
|||||
1277 GCTCGCGGTTTAACGGGTGATCAGACTCCCGCTCTGCGTGATTTCTCTAG 1326
94 nGlyLysVal.....
:
1327 CGATAAACCTGCACAAAAATATTATTTTCGTGATTCGGCATGGGATGGGG 1376
98 ..PheAspValThrLysGlyArgLysPheTyrGlyProGluGlyProTyr 113
```



OM of: US-09-203-548-1 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Apr 19, 2000 3:53 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09203548/runat\_14042000\_104704\_23251/app\_query.fasta.1  
-DB=Issued\_Patents\_NA -QMT=fastcap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=100000 -USER=US09203548 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

Search information block:

Query: US-09-203-548-1  
Query length: 220  
Database: Issued\_Patents\_NA:\*  
Database sequences: 214294  
Database length: 59861574  
Search time (sec): 23.300000

seq\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-960-022-5		1061.00	2394.76	5.9e-126	1868
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-167-035-9		81.00	145.69	1.12	2345
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-208-887A-9		81.00	145.69	1.12	2345
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-539-005-9		81.00	145.69	1.12	2345
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-952-817-8		80.50	143.17	1.54	2639
/cgn2_6/ptodata/1/ina/backfiles1.seq:5210025-1		80.50	143.17	1.54	2639
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-906-349A-7		77.00	136.52	3.62	2345
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-890-094-3		77.00	136.15	3.79	2420
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-960-022-9		76.00	128.80	9.74	3742
/cgn2_6/ptodata/1/ina/5D_COMB.seq:PCT-US91-09055-1		75.00	117.14	43.44	8378
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-814-964-9		74.50	136.42	3.66	1444
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-258-442-9		74.50	136.42	3.66	1444
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-328-809-4		74.50	136.42	3.66	1444
/cgn2_6/ptodata/1/ina/5A_COMB.seq:PCT-US92-11107-9		74.50	136.42	3.66	1444
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-814-964-8		74.50	133.24	5.51	1898
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-258-442-8		74.50	133.24	5.51	1898
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-328-809-3		74.50	133.24	5.51	1898
/cgn2_6/ptodata/1/ina/5D_COMB.seq:PCT-US92-11107-8		74.50	133.24	5.51	1898
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US91-04541-1		74.50	129.91	8.44	2529
/cgn2_6/ptodata/1/ina/backfiles1.seq:5262319-1		74.50	129.91	8.44	2529
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-814-964-6		74.50	128.57	10.03	2839
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-258-442-6		74.50	128.57	10.03	2839
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-328-809-1		74.50	128.57	10.03	2839
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US92-11107-6		74.50	128.57	10.03	2839
/cgn2_6/ptodata/1/ina/backfiles1.seq:5168051-10		73.50	129.38	9.04	2173
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-890-094-17		72.50	127.65	11.28	2070
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-068-109-1		72.00	132.16	6.32	1272
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US92-05532-1		72.00	127.44	11.59	1910
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-272-882D-1		72.00	125.48	14.90	2261
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-484-128-6		71.50	137.02	3.39	759
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-833-610-1		71.00	126.09	10.67	1483
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-248-1		71.00	116.09	49.69	4165
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-440-815-1		71.00	116.09	49.69	4165
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-845-998-1		70.50	126.20	13.59	1581
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-211-930-12		70.50	123.31	19.70	2028
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-211-930-1		70.00	116.86	45.04	3201
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-117-491-25		70.00	116.09	49.71	3420
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-271-364A-6		70.00	116.09	49.71	3420
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-222-715B-25		70.00	116.09	49.71	3420
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-004-1398-2		70.00	112.38	80.01	4707
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-492-2		70.00	112.38	80.01	4707

/cgn2\_6/ptodata/1/ina/PCTUS9\_COMB.seq:PCT-US96-10545A-2 + 70.00 112.38 80.01  
/cgn2\_6/ptodata/1/ina/PCTUS9\_COMB.seq:PCT-US96-05611A-12 - 70.00 105.45 194.38  
/cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-08-853-310-3 + 69.00 108.57 130.36  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-463-911-1 + 68.50 124.10 17.78

seq\_name: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-08-960-022-5

seq\_documentation\_block:

Sequence 5: Application US/08960022  
Patent No. 5978837  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: LaVallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,022  
FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1868 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-960-022-5

alignment\_scores:

Quality: 1061.00 Length: 222  
Ratio: 5.005 Gaps: 2  
Percent Similarity: 95.495 Percent Identity: 94.595

alignment\_block:

US-09-203-548-1 x US-08-960-022-5

Align seg 1/1 to: US-08-960-022-5 from: 1 to: 1868

- 1 MetAlaAGluAspValValAlaThrGlyAlaAspProSerAspLeuCl 17
- 52 ATGGCTCGCGAGGATGTGGTGGCGACTGGCGCCGACCAAGCATCTGGA 101
- 17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuL 34
- 102 GAGCGCGCGGCTGCTGCATGAGATTTTACGTCGCGCGCTCAACCTGCTGC 151
- 34 euLeuGlyLeuCysIlePheLeuLeuTyrlLysIleValArgGlyAspGln 50

```
|||||
152 TGCTTGGCTCTGCATCTTCCGCTCTACAAGATCGTGGCGGGACCAG 201
|||||
51 ProAlaIaSerGlyAspArgThrThr***ProProProLeuProAr 67
|||||
202 CGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 251
|||||
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGly 84
|||||
252 CCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301
|||||
84 alGlnAspProArgLeuLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
302 TCCAGAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 351
|||||
101 ThrLysGlyArgLysPheThrGlyProGluGlyProThrGlyValPheAl 117
|||||
352 ACCAAAGCGCGCAATTTTACSGGGCGGCGGCGGCGGCGGCGGCG 401
|||||
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGlu 134
|||||
402 TGAAGAGATGATCCAGGGGCGGCGGCGGCGGCGGCGGCGGCGG 451
|||||
134 laLeuLysAspGluThrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
452 CACTCAAGGATGAGTACGATGACCTTCTGACCTCACTGCTGCCACG 501
|||||
151 GluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyHisHisVa 167
|||||
502 GAGACTCTGAGTGACTGGGAGCTCAGTTCACTTCAAGTATCATCACGT 551
|||||
167 lGlyLysLeuLeuLysGluGlyGluProThrValTySerAspGlu 184
|||||
552 GGGCAACTCTGTAAGGAGGGGAGGAGCCACTGTGTACTCAGATGAGG 601
|||||
184 luGluProLysAspGluSerArgLysAsn.ValLysAlaPheSerGln 200
|||||
602 AAGAACCAAAAGATGAGAGTGCGCGGAAAAATGATTAAAGCATTCAGTG 651
|||||
200 ySerIleSer.*****TyPheAlaLysSerPheValThrValHis*** 216
|||||
652 AAGTATATCTATTTTGTATTTTGAATAATCATTTGTAACAGTCCACTCT 701
|||||
217 ValPheLysThr 220
|||||
702 GTCTTTAAACA 713
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-167-035-9

```
seq_documentation_block:
; Sequence 9, Application US/08167035
; Patent No. 5618691
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Skolnick, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: 10036-2711
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/167.035
; FILING DATE: 16-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2345 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 369..1973
US-08-167-035-9

alignment_scores:
Quality: 81.00 Length: 85
Ratio: 1.800 Gaps: 3
Percent Similarity: 52.941 Percent Identity: 27.059

alignment_block:
US-09-203-548-1 x US-08-167-035-9 ..
Align seg 1/1 to: US-08-167-035-9 from: 1 to: 2345

85 GlnAspProArgIleLeuMetAlaIle.....AsnGlyLysVa 97
:::|||||::: ||::: ||:
1176 AGGACCCCGACACACCTACAGTATGCGCAGATGTGAATGAGTCCAATCT 1225

97 lPheAspValThrLysGlyArgLysPheThrGlyProGluGlyProTyrg 114
::: |||||::: |||||::: |||||
1226 CTATGTGTGACCCAGGCGCGCAAGCTGTGGATGCCCACTGACTTCG 1275

114 lYValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeu 130
||
1276 GC.....TTCTGTGTC 1286

131 AspLysGluAlaLeuLysAspGluTyArgAspLeuSerAspLeuThrAl 147
::: |||||::: ||: ||:
1287 AGCCCAACAGCTTCGAACCGCCACAGGGGCTCCACATCTTCTGCGAC 1336

147 aAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPhe...ThrPheL 163
: |||||::: ||: ||: ||: |||
1337 TGAGGATGACGAGACTCGGACCTGCTGGCTGCTGCTCCGCTCTTCA 1386

163 ystyr 164
|||||
1387 AGTAC 1391
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-208-887A-9

```
seq_documentation_block:
; Sequence 9, Application US/08208887A
; Patent No. 5677421
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Skolnick, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
```



```

; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: 10036-2711
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,887A
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2345 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 369...1973
; US-08-208-887A-9

alignment_scores:
    Quality: 81.00    Length: 85
    Ratio: 1.800      Gaps: 3
Percent Similarity: 52.941    Percent Identity: 27.059

alignment_block:
US-09-203-548-1 x US-08-208-887A-9 ..

Align seg 1/1 to: US-08-208-887A-9 from: 1 to: 2345

85 GlnAspProArgIleLeuMetAlaIle.....AsnGlyLysVa 97
   :::::::::::::: :::::
1176 AAGCACCACACACCTACAGTATGTGGCAGATGTGAATGAGTCCAATGT 1225

97 lPheAspValThrLysGlyArgLysPheTyrgGlyProGluGlyProTyrg 114
   :::::::::::::: :::::
1226 CTATGTGTGACCCAGCGCCGCAAGCTGTATGGGATGCCACTGACTTCG 1275

114 lValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeu 130
   :::::::::::::: :::::
1276 GC.....TTCGTGTC 1286

131 AspLysGluAlaLeuLysAspGluTyArgAspLeuSerAspLeuThrAl 147
   :::::::::::::: :::::
1287 AAGCCCAACAGCTTCGAAGCCGACAGAGGGCTCCACATCTTCGACG 1336

147 aAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPhe...ThrPheL 163
   :::::::::::::: :::::
1337 TGAGGATGAGCAGACTCGGACCTGCTGCTGCTGCTGCTGCTGCTTCA 1386

163 yTyr 164
   ::::::::::::::
1387 AGTAC 1391

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq.US-08-539-005-9
seq_documentation_block:
; Sequence 9, Application US/08539005

```

```

; Patent No. 5858686
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Joseph
; APPLICANT: Skolnick, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: 10036-2711
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,005
; FILING DATE: 4-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,035
; FILING DATE: 16-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2345 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 369...1973
; US-08-539-005-9

```

```

alignment_scores:
    Quality: 81.00    Length: 85
    Ratio: 1.800      Gaps: 3
Percent Similarity: 52.941    Percent Identity: 27.059

alignment_block:
US-09-203-548-1 x US-08-539-005-9 ..

Align seg 1/1 to: US-08-539-005-9 from: 1 to: 2345

85 GlnAspProArgIleLeuMetAlaIle.....AsnGlyLysVa 97
   :::::::::::::: :::::
1176 AAGCACCACACACCTACAGTATGTGGCAGATGTGAATGAGTCCAATGT 1225

97 lPheAspValThrLysGlyArgLysPheTyrgGlyProGluGlyProTyrg 114
   :::::::::::::: :::::
1226 CTATGTGTGACCCAGCGCCGCAAGCTGTATGGGATGCCACTGACTTCG 1275

114 lValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeu 130
   :::::::::::::: :::::
1276 GC.....TTCGTGTC 1286

131 AspLysGluAlaLeuLysAspGluTyArgAspLeuSerAspLeuThrAl 147

```





1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

```

alignment_scores:
  Quality: 76.00      Length: 351
  Ratio: 0.613       Gaps: 18
  Percent Similarity: 35.328  Percent Identity: 21.368

alignment_block:
US-09-203-548-1 x US-08-960-022-9 ..

Align seg 1/1 to: US-08-960-022-9 from: 1 to: 3742

2 AlaAlaGluAspValValAlaIaThrGly.....AlaAspProSerAs 15
   ::::::::::::::::::::|||
1043 TCAACCCAGAGCGTCTGCCGATGGGACAGATCTGCAGACACCCCTCACC 1092
   ::::::::::::::::::::|||

15 pLeuGluSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnL 32
   ::::::::::|||
1093 AGTCCACAAGATGGG...CAGAATGAGCCGACAGTGCACCAAGAAGACC 1139
   ::::::::::::::::::::|||

32 euLeuLeuLeuGlyLeuCysIlePheLeuLeuTyrlsIleValaArgGly 48
   ::::::::::|||
1140 TCCACTGTGTGGGGACCAGC...AGGCTGCTCTATCACATCACTGATGTT 1186
   ::::::::::::::::::::|||

49 AspGlnProAlaIaAspSerGly..... 55
1187 GATAACCCACTGCTCTCGCCAGCATGCTCCACTTTCAGCCAAAGCCAGAG 1236
   ::::::::::|||

56 .....AspArgThrThr***ProProPro..... 64
   ||| ::::::::::
1237 ATTCAACTTAGACCCCGAGTCAGCCCATCTCCACCCACGACTCAGCAGT 1286
   ::::::::::

65 .....LeuProArg..... 67
   ::::::::::

1287 TTATGATCCGCGGAGTCTTTCACGCTCGAGCTGTGGAGATGCCAGGAG 1336
   ::::::::::

68 .....LeuLysArgAr 71
   ||| ::::::::::

1337 CCACAGACCATCACCCAGCTCACCAGACACATCCAGAGAGCTCAAGCGGAA 1386
   ::::::::::

71 gAspPheThrProAlaGluLeuArgArPhe..... 81
   ::::::::::|||

1387 A.....ATTCGGAATTTGCAGAAAAATTTGAACAAG 1418
   ::::::::::|||

82 .....AspGlyValGlnAspProArgIle 89
   ::::::::::|||

1419 AAAAGAAATACGGCTTTCATGTTGTGACAGACTTCTAATCTGAAGTTC 1468
   ::::::::::|||

```

```
90  LeuMetAlaIleAsnGlyLysValPheAspValThrLysGlyArgLysPh 106
    |||  ::|||  |||:::|||||
1469  CTGAATGGATGAAT.....GATTGGCTAAAGTGTCTAAACA 1506
    |||  ::|||  |||:::|||||
106  eTyr.....GlyPro 110
    |||  ::|||  |||:::|||||
1507  GCTCAAGAAGTAAGCTGTCAGAGAACAAGGAGTCTCCCA 1556
    |||  ::|||  |||:::|||||
110  LuGlyPro..... 112
    |||  ::|||  |||:::|||||
1557  AAGTCCACCTAGAAACCTGTTGTGAGCAACCCACAGTCCCAAGAA 1606
    |||  ::|||  |||:::|||||
113  .....TyrGlyValPheAlaGlyArgAspAlaSerArgGly..... 124
    |||  ::|||  |||:::|||||
1607  AATGGAAACCGAAGCTGGCGCGCGGCAAGCTCTCTGGAGAAGA 1656
    |||  ::|||  |||:::|||||
125  .....LeuAlaThrPheCysLeu..... 130
    |||  ::|||  |||:::|||||
1657  GACTCCAGATGCTGCCTTGACATGCCTGAAGAGAGAGAGCAACTTC 1706
    |||  ::|||  |||:::|||||
131  .....AspLysGluAlaLeuLys 136
    |||  ::|||  |||:::|||||
1707  CTCCCGAGGAGATTCTAGGTAACTAAGCAAGACAAGAACCTCATAAG 1756
    |||  ::|||  |||:::|||||
137  AspGluTyrAsp..... 140
    |||  ::|||  |||:::|||||
1757  CGCTTTATGACCGCATACAGATTATCAAGCAAAATCTGTCAACACCTTC 1806
    |||  ::|||  |||:::|||||
141  .....AspLeuSerAspLeu 146
    |||  ::|||  |||:::|||||
1807  CTTATTCCAACTTACAGAGGAGGACTCTGTAGAGAGAGAGAGCACTCC 1856
    |||  ::|||  |||:::|||||
146  hrAlaAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPheThrPhe 162
    |||  ::|||  |||:::|||||
1857  AGGAAGCAACCAACCTCTTTGGCAGATCCAGCATCTCAC..... 1897
    |||  ::|||  |||:::|||||
163  LysTyrHisHisValGlyLysLeuLysGluGlyGluGluProThrVa 179
    |||  ::|||  |||:::|||||
1898  .....CTTCCTGTTGGTGACCACTCAC.. 1921
    |||  ::|||  |||:::|||||
179  LTrpSerAspGluGluGluPro.....LysAspGluSers 191
    |||  ::|||  |||:::|||||
1922  TACTCTAATGACACTGAGCTGTAGGGCCCTTTTACCAAGATGAA... 1966
    |||  ::|||  |||:::|||||
191  exArgLysAsnValLysAlaPheSerGlySerIleSer****TyrPhe 207
    |||  ::|||  |||:::|||||
1967  ..AAGAAAGAAAGTAAACCAACCACTCTCTCCATGTCTAATTTACATGAG 2014
    |||  ::|||  |||:::|||||
208  Ala 208
    |||  ::|||  |||:::|||||
2015  GCT 2017
```

seq\_name: /cgn2\_6/ptodata/1/lna/PCTUS9\_COMB.seq:PCT-US91-09055-1

seq\_documentation\_block:

```
; Sequence 1, Application PC/TUS9109055
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan Marc and Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: Purified SLIT protein and Sequence Elements Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University
; ADDRESS: Office of Cooperative Research
; STREET: 246 Church Street
; STREET: Suite 401
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
; COMPUTER: Apple Macintosh
```

```
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09055
; FILING DATE: 19911127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/624,135
; FILING DATE: 7-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard J.
; REGISTRATION NUMBER: 28,180
; REFERENCE/DOCKET NUMBER: 900964/RSB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 972-1400
; TELEFAX: (212) 370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8378
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Drosophila Melanogaster
; DEVELOPMENTAL STAGE: embryonic and larval, germ-line
; IMMEDIATE SOURCE:
; LIBRARY: cDNA and Genomic
; CLONE: be2.4, ka2.4, B52-2, B52-5, smart2-19
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 2R
; MAP POSITION: 52D
; UNITS: chromosome band
; FEATURE:
; NAME/KEY: 5' leader sequence
; LOCATION: 1 to 314
; IDENTIFICATION METHOD: experimental
; NAME/KEY: Translated region
; LOCATION: 315 to 4754
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: codes for slit protein
; NAME/KEY: 3' untranslated region
; LOCATION: 4755 to 8378
; IDENTIFICATION METHOD: experimental
; PCT-US91-09055-1
```

alignment\_scores:

```
Quality: 75.00 Length: 143
Ratio: 1.042 Gaps: 9
Percent Similarity: 50.350 Percent Identity: 28.671
```

alignment\_block:

US-09-203-548-1 x PCT-US91-09055-1 ..

Align seg 1/1 to: PCT-US91-09055-1 from: 1 to: 8378

```
82 AspGlyValGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPh 98
    |||  ::|||  |||:::|||||
5445  GACGCGCTCGATGACGAA.....TTCGCCGACACGGGGGAG..... 5480
    |||  ::|||  |||:::|||||
98  eAspValThrLysGlyArgLysPheTyrGlyProGlu..... 110
    |||  ::|||  |||:::|||||
5481  .....AACGAGGGCGCGGCTTTTGGCTCCCGACGACGACAGCGCA 5523
    |||  ::|||  |||:::|||||
111  .....GlyProTyrGlyValPheAlaGlyArgAspAlaSerArgGlyLeu 125
    |||  ::|||  |||:::|||||
5524  AGAAGGTCTCGTATCATCCCGCAAGAACGACGATCCCAATCATC 5573
    |||  ::|||  |||:::|||||
126  AlaThrPheCysLeuAspLysGluAlaLeu.....LysAspGluTy 139
```

```

5574 TCCAGC...CGCTGGCAAGGTGACGATTGTGTACCAGCAGCAGCAAA 5620
139 rAspAspLeuSerAspLeuThrAlaAlaGlnGlu.....T 152
5621 GGACCAAGTCGGGACCAAGGATCGCGCAACAGCAGCAGCAAAAGCCGG 5670
152 hrLeuSerAspTrp.....GluserglnPheThr 161
5671 CGCTCAGCAGATTTCAGCCCTGTGTACCGGACCCGAGAGCAGTATCGC 5720
162 PheLysThrHisHis.....ValGlyLysLeuLeuLysGluGlyG1 175
5721 TTTCCGTCGCCCCACCCCAAGATTACACTGTTCTGACGCCCGATGGCAA 5770
175 uGluProThrValThr....SerAspGluGluGluProLysAspGluSerS 191
5771 GGTGGCGTGTCTATTCGCGGAGACTCGGAGACTCCAAGTACGAGCCCA 5820
191 erArgLysAsnValLysAlaPheSerGly 200
5821 TACGCAACTGACGACCAAGTTTTCGGGA 5849
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-07-814-964-9

```

seq_documentation_block:
: Sequence 9, Application US/07814964
: Patent No. 5359047
: GENERAL INFORMATION:
: APPLICANT: Donahue, Brian A.
: APPLICANT: Toney, Jeffrey H.
: APPLICANT: Bruhn, Suzanne L.
: APPLICANT: Pil, Pieter M.
: APPLICANT: Brown, Steven
: APPLICANT: Kellelt, Patti
: APPLICANT: Essigmann, John M.
: APPLICANT: Lippard, Stephen J.
: TITLE OF INVENTION: DNA Structure Specific Recognition
: TITLE OF INVENTION: Protein and Uses Therefor
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: 2 Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/814,964
: FILING DATE: 19911226
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/539,906
: FILING DATE: 18-JUN-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-4787AAA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1444 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
```

```

: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: LIBRARY: Human B cell
: CLONE: lambda Pt2
: US-07-814-964-9

alignment_scores:
  Quality: 74.50      Length: 206
  Ratio: 0.753       Gaps: 12
  Percent Similarity: 48.058  Percent Identity: 22.330

alignment_block:
US-09-203-548-1 x US-07-814-964-9 ..
Align seg 1/1 to: US-07-814-964-9 from: 1 to: 1444

49  AspGlnProAlaAlaSerGlyAsp...ArgThrThrThr***ProProPr 64
|||||
324 GATCAGCTGGATCCCAATCAAGCAAGGCCAAACTCGCTACCACCTCC 373

64  oLeuProArgLeu.....LysArgArgAspPheThrProAlaGlu.Leu 78
|||
374 TGATCCTCCTCTTCTCCAAGGACGAGGACATTTCGTGACTCTGAACATG 423

79  .....ArgArgPheAspGlyValGlnAspProArgIl 89
|||||
424 AACGAGGAAGAGTGGAGAAGCGCTTTGAGGT.....CGGCT 461

89  eLeuMetalIleAsnGlyLysValPheAsp..... 99
|||||
462 CACCAAGAACATGTGAGATCCCTCTATGAGATGTCAGCCGGTCATGA 511

100 .....ValThrLysGlyArgLysPheThrGlyPro 109
|||||
512 AAGCACTGTAAACCGCAAGATCACAGTGCCAGGCAACTTCCAAGGGCAC 561

110 GluGlyProThrGlyValPheAlaGlyArgAspAlaSerArgGlyLeu.. 125
|||||
562 TCAGGGGCCAGTGCACTTACCTGTCTCTACAAGCAAGCTCAGGACTGCT 611

126 .....AlaThrPheCysLeuAspLysGluAlaLeuLysA 137

612 CTACCCGCTGGAGCGGGCTTCATCTACGTCACCAAGCCACCTGTGCACA 661

137 spGluThrAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153
|||||
662 TCCGCTTCGATGAGATCTCCTTTGTCAACTTGTCTCGTGGTACCACACT 711

154 Ser.....AspTrpGlu.....SerGlnPheTh 161
|||
712 ACTCGTTCCTTTGACTTTGAAATTGAGACCAAGCAGGCGACTCAGTATAC 761

161 rPhe.....LysThrHisHisValGlyLysLeu..... 170
|||||
762 CTTACAGCAGCATTTGAGAGGGAGGAGTACGGGAACCTGTTGATTTGTCA 811

171 .....LeuLysGluGlyGluGlu 176
|||||
812 AGCGGAAAAAGCTCAACATCAAAACCAGGAGATTGAAGAGGGCATGAC 861

177 ProThrVal.....TyrSerAspGluGluGluProLysAspGluSerS 191
|||||
862 CCAAGCTACGATGAATATGCTGACTCTGTATGAGGACCAGCATGATGCCTA 911

191 rArgLysAsnValLys 196
|||||
912 CTTGGAGAGGATGAAG 927

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-258-442-9
seq_documentation_block:
```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328.809
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Penton, Gallian M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-248-7000
; TELEFAX: 617-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1444 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: Human B cell
; CLONE: lambda Pt2
; US-08-328-809-4

```

```

alignment_scores:
  Quality: 74.50      Length: 206
  Ratio: 0.753        Gaps: 12
Percent Similarity: 48.058 Percent Identity: 22.330

```

## alignment\_block:

```

US-09-203-548-1 x US-08-328-809-4
Align seg 1/1 to: US-08-328-809-4 from: 1 to: 1444

49 AspGlnProAlaAlaSerGlyAsp...ArgThrThrThr***ProProPr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 GATCAGCTGGATCCCAATCAAGCAAGCCAACTCGTACCATTCC 373
64 oLeuProArgLeu.....LysArgArgAspPheThrProAlaGlu..Leu 78
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 TGATCCTCTCTCTCCCAAGGACGAGACATTCGTGTGACTCTGAACATG 423
79 .....ArgArgPheAspGlyValGlnAspProArg1l 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 AACGAGGAAGAGTGGAGGAGCGCTTGAGGT.....CGGCT 461
89 eLeuMetAlaIleAsnGlyLysValPheAsp..... 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
462 CACCAAGAACATGTCAGGATCCCTCTATGAGATGGTCAGCCGGTCATGA 511
100 .....ValThrLysGlyArgLysPheThrGlyPro 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
512 AAGCACTGGTAACCGCAAGATCACAGTCCAGGCAACTTCCAAGGGCAC 561
110 GluGlyProTyrGlyValPheAlaGlyArgAspAlaSerArgGlyLeu.. 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
562 TCAGGGGCCAGTGCATCCCTGTTCTCTCAAGGCAAGCTCAGGACTGCT 611
126 .....AlaThrPheCysLeuAspLysGluAlaLeuLysA 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
612 CTACCCGCTGGAGCGGGCTTCATCTACGTCACAGCCACCTGTGCACA 561
137 spGluTyrAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
662 TCCGCTTCGATGAGATCCTCTTGTGCAACTTTCGTCGTGTACCACTACT 711
154 Ser.....AspTrpGlu.....SerGlnPheTh 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
712 ACTCGTTCCTTTGACTTTGAAATTGAGACCAAGGCGCACTCAGTATAC 761
161 rPhe.....LysTyrHisHisValGlyLysLeu..... 170

```

alignment\_scores:

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
762 CTTACGACGATTGAGAGGAGGAGTACGGAACTGTTTGATTTGTCA 811
171 .....LeuLysGluGlyGluGlu 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
812 ACAGGAAAAGTCAACATCAAAACCCGAGGATTGAAGAGGGCATGAAC 861
177 ProThrVal.....TyrSerAspGluGluProLysAspGluSerSe 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
862 CCAAGCTACGATCAATATGCTGACTCTGATGAGGACCATCATGCCTA 911
191 rArgLysAsnValLys 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
912 CTTGAGAGGATGAAG 927

seq_name: /cgn2_6/ptodata/1/ina/PCUS9_COMB.seq:PCT-US92-11107-9

seq_documentation_block:
; Sequence 9, Application PC/TUS9211107
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11107
; FILING DATE: 19921218
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1444 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: Human B cell
; CLONE: lambda Pt2
; PCT-US92-11107-9

```



```
Quality: 74.50 Length: 206
Ratio: 0.753 Gaps: 12
Percent Similarity: 48.058 Percent Identity: 22.330

alignment_block:
US-09-203-548-1 x PCT-US92-11107-9 ..
Align seg 1/1 to: PCT-US92-11107-9 from: 1 to: 1444
49 AspGlnProAlaAlaSerGlyAsp...ArgThrThrThr***ProProPr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 GATCAGCCTGGATCCCAATCAAGCAAGGCCAACTCGCTACCACTTCC 373
64 oLeuProArgLeu.....LysArgArgAspPheThrProAlaGlu.Leu 78
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 TGATCCTCTCTCTCCCAAGGACGAGGACATTTGCTGACTCTGACATG 423
79 .....ArgArgPheAspGlyValGlnAspProArgI 89
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 AACGAGGAAGAGTGGAGAGCGCTTTGAGGT.....CGGCT 461
89 eLeuMetalalleAsnGlyLysValPheAsp..... 99
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 CACCAAGAACATGTCAGGATCCCTCTATGAGATGTCAGCGGGTCAATGA 511
100 .....ValThrLysGlyArgLysPheThrGlyPro 109
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
512 AAGCACTGTTAAACCGCAAGATCAGTCGCGGCACTTCAAGGGCAC 561
110 GluGlyProTyrGlyValPheAlaGlyArgAspAlaSerArgGlyLeu.. 125
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
562 TCAGGGCCCGAGTCATCTACTCTGTTCTTACAAAGCAAGCTCAGGACTCT 611
126 .....AlaThrPheCysLeuAspLysGluAlaLeuLysA 137
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
612 CTACCCGCTGGAGGGGGCTTCATCTACGTCACCAAGCCACCTGTGCACA 661
137 spGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
562 TCCGCTTCGATGAGATCTCTCTTCTCAACTTGTCTGTTGGTGGTACCAC 711
154 Ser.....AspTrpGlu.....SerGlnPheTh 161
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
712 ACTCGTTCTTGTACTTTGAATTTGAGACCAAGCAGGCGCATCATGATAC 761
161 rPhe.....LysTyrHisHisValGlyLysLeu..... 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
762 CTTCAGCAGCATTCAGAGGGAGGAGTACGGGAAACTGTTTGTATTGTCA 811
171 .....LeuLysGluGlyGluGlu 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
812 ACGGGAAAAAGCTCAACATCAAAAACCGAGGATTGAAAGAGGGCATGAAC 861
177 ProThrVal.....TyrSerAspGluGluProLysAspGluSerSe 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
862 CCAAGCTACGATGATGATGCTGATCTGATGATGAGGACCATGATGCTTA 911
191 rArgLysAsnValLys 196
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
912 CTTGGAGAGGATGAAG 927
```

seq\_name: /cgn2\_6/ptodata/1/Ina/5A\_COMB.seq:US-07-814-964-8

```
seq_documentation_block:
; Sequence 8, Application US/07814964
; Patent No. 5359047
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
```

```
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: Human B cell
; CLONE: lambda-Pt1
; US-07-814-964-8
```

```
alignment_scores:
Quality: 74.50 Length: 206
Ratio: 0.753 Gaps: 12
Percent Similarity: 48.058 Percent Identity: 22.330

alignment_block:
US-09-203-548-1 x US-07-814-964-8 ..
Align seg 1/1 to: US-07-814-964-8 from: 1 to: 1898
```

```
49 AspGlnProAlaAlaSerGlyAsp...ArgThrThrThr***ProProPr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 GATCAGCCTGGATCCCAATCAAGCAAGGCCAACTCGCTACCACTTCC 373
64 oLeuProArgLeu.....LysArgArgAspPheThrProAlaGlu.Leu 78
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 TGATCCTCTCTCTCCCAAGGACGAGGACATTTGCTGACTCTGACATG 423
79 .....ArgArgPheAspGlyValGlnAspProArgI 89
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 AACGAGGAAGAGTGGAGAGCGCTTTGAGGT.....CGGCT 461
89 eLeuMetalalleAsnGlyLysValPheAsp..... 99
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 CACCAAGAACATGTCAGGATCCCTCTATGAGATGTCAGCGGGTCAATGA 511
100 .....ValThrLysGlyArgLysPheThrGlyPro 109
```

```

512  AAGCACTGGTAAACCGCAAGATCACAGTCGAGGCAACTTCCAAGGGCAC 561
      ::::| | | | | | | | | |
110  GluGlyProTyrGlyValPheAlaGlyArgAspAlaSerAArgGlyLeu... 125
      ::||| ::::| | | | | | | | | |
562  TCAGGGCCCCAGTCGATACCTTGTTCTTACAAGCAAGCTCAGGACTGCT 611
      ::||| ::::| | | | | | | | | |
126  .....AlaThrPheCysLeuAspLysGluAlaLeuLysA 137
      ::::| | | | | | | | | |
612  CTACCCGCTGGAGGGGGCTTCATCTACGTCACCAAGCAAGCTGTGCACA 661
      ::::| | | | | | | | | |
137  spGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnGlnGluThrLeu 153
      ::::| | | | | | | | | |
662  TCCGCTTCGATGAGATCCTCTTGTCAACTTGTCTGGTGATACCACACT 711
      ::::| | | | | | | | | |
154  Ser.....AspTrpGlu.....SerGlnPheTh 161
      ::| | | | | | | | | |
712  ACTCGTTCCITTGACTTTGAATTCAGACCAGGAGGCACTCAGTATAC 761
      ::::| | | | | | | | | |
161  rPhe.....LysTyrHisHisValGlyLysLeu..... 170
      ::::| | | | | | | | | |
762  CTTCAGCAGCATTGAGAGGGAGGAGTACGGGAAACTGTTGATTTTGTCA 811
      ::::| | | | | | | | | |
171  .....LeuLysGluGlyGluGlu 176
      ::::| | | | | | | | | |
812  ACGCGAAAAAGCTCAACATCAAAAAACCGAGGATTGAAAGAGGGCATGAAC 861
      ::::| | | | | | | | | |
177  ProThrVal.....TyrSerAspGluGluGluProLysAspGluSerSe 191
      ::||| ::::| | | | | | | | | |
862  CCAAGCTACGATGAATATGCTGACTCTGATCAGGACCACGACATGATGCCTA 911
      ::||| ::::| | | | | | | | | |
191  rArgLysAsnValLys 196
      ::::| | | | | | | | | |
912  CTTGGGAGGATGAAG 927
      ::::| | | | | | | | | |

```

OM of: US-09-203-548-1 to: EST:\* out\_format : pfs

Date: Apr 19, 2000 3:26 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_p2n\_model -DEV=xlp  
-Q/cgn2\_l/USPto.spool/US09203548/runat\_14042000\_104704\_23231/app\_query.fasta.1  
-DB=EST -QMT=fastcap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.500 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09203548  
-NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-203-548-1

Query length: 220

Database: EST:\*

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 270.200000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est28:AI479776	-	902.00	1842.47	7.3e-94	720	AI479776 tm69c04.x1 NCI_CGAP_Br
gb_est26:AI394728	-	745.00	1521.10	5.8e-76	567	AI394728 tg24g02.x1 NCI_CGAP_C
gb_est20:AA048529	-	729.00	1488.72	3.7e-74	534	AA048529 EST191289 Normalized
gb_est10:AA184910	-	720.00	1470.14	4.0e-73	535	AA184910 mt62c03.r1 Soares 2NM
gb_est9:AA117455	-	709.00	1447.90	7.0e-72	512	AA117455 mn28g11.r1 Beddington
gb_est8:AA023630	-	697.00	1427.38	1.6e-70	501	AA023630 mh80f05.r1 Soares mous
gb_est36:AI1885620	-	687.50	1361.11	4.8e-67	573	AI1885620 w159e09.x1 NCI_CGAP_Br
gb_est38:AW018545	-	641.00	1307.43	4.7e-64	522	AW018545 fd48c10.y1 zebrafish w
gb_est28:AI477883	-	640.00	1305.42	6.0e-64	519	AI477883 fb57b05.y1 zebrafish w
gb_est29:AI575080	-	639.00	1304.53	6.7e-64	464	AI575080 UI-R-GO-ul-b-08-0-UI-s
gb_est6:N52291	-	636.00	1298.48	1.5e-63	458	N52291 yv48a12.s1 Soares fetal
gb_est34:AI1794380	-	636.00	1296.32	1.9e-63	563	AI1794380 fc45g04.y1 zebrafish w
gb_est12:AA288129	-	630.00	1285.02	8.2e-63	508	AA288129 vb12e11.r1 Soares mous
gb_est27:AI453643	-	625.00	1274.36	3.2e-62	525	AI453643 tj45h05.x1 Soares NSF
gb_est14:AA428520	-	621.50	1269.08	6.4e-62	531	AA428520 zw47c07.r1 Soares tota
gb_est11:AA274818	-	621.00	1266.05	9.4e-62	528	AA274818 vb02b12.r1 Soares mous
gb_est7:W56474	-	615.50	1254.86	3.9e-61	520	W56474 zc59b01.r1 Soares parath
gb_est38:AL035776	-	586.00	1236.15	4.3e-60	478	AL035776 DKFp564M0882.r1 564 (
gb_est26:AI334635	-	579.00	1197.26	6.4e-58	464	AI334635 tb20f05.x1 NCI_CGAP_K
gb_est14:AA389446	-	579.00	1181.37	4.9e-57	438	AA389446 mp19e12.r1 Life Tech m
gb_est9:AA081900	-	574.00	1169.64	2.2e-56	501	AA081900 zn23g04.r1 Stratagene
gb_est9:AA111285	-	560.00	1141.30	8.3e-55	476	AA111285 mo53h02.r1 Life Tech m
gb_est16:AA572534	-	550.00	1121.76	1.0e-53	429	AA572534 v184d11.r1 Stratagene
gb_est9:AA101294	-	533.00	1086.69	9.2e-52	429	AA101294 zn71f03.r1 Stratagene
gb_est18:AA699865	-	531.50	1083.60	1.4e-51	429	AA699865 zj81e01.s1 Soares fetal
gb_est9:AA106719	-	526.00	1074.33	4.5e-51	352	AA106719 mm19h05.r1 Stratagene
gb_est5:H95223	-	516.50	1053.10	6.8e-50	411	H95223 yv59f10.r1 Soares placen
gb_est9:AA115422	-	507.00	1033.63	8.3e-49	406	AA115422 z18se09.r1 Stratagene
gb_est40:AA133364	-	498.50	1011.11	1.5e-47	653	AA133364 f119g11.y1 Sugano kawa
gb_est10:AA166645	-	495.00	1009.30	1.9e-47	390	AA166645 zq39b04.r1 Stratagene
gb_est26:AI332276	-	494.00	1006.76	2.6e-47	408	AI332276 fa97f10.y1 zebrafish f
gb_est6:N56942	-	489.00	996.37	9.9e-47	411	N56942 za48c12.s1 Soares fetal
gb_est5:N8287	-	481.50	982.77	5.8e-46	344	N8287 yv68c06.s1 Soares fetal
gb_est9:AA088127	-	477.00	972.41	2.1e-45	381	AA088127 mm9a10.r1 Stratagene
gb_est10:AA147865	-	475.50	963.99	6.3e-45	633	AA147865 zo47a10.r1 Stratagene
gb_est11:Z84149	-	470.50	957.35	1.5e-44	446	Z84149 SS284149 Porcine small i
gb_est9:AA080939	-	470.00	959.48	1.1e-44	330	AA080939 zn18a05.r1 Stratagene
gb_est25:AI317729	-	469.00	957.17	1.5e-44	338	AI317729 u16h09.y1 Sugano mous
gb_est8:AA016712	-	463.00	941.86	1.1e-43	367	AA016712 mg90h03.r1 Soares mous
gb_est39:AA125087	-	455.00	931.60	4.0e-43	300	AA125087 UI-M-BHJ.1-afx-f-04-0-
gb_est7:W97563	-	450.00	926.49	7.7e-43	401	W97563 mg02a02.r1 Soares mous
gb_est35:AI843194	-	454.00	927.47	6.8e-43	300	AI843194 UI-M-AKI-aet-d-12-0-UI

gb\_est3:R59281 + 454.00 925.30 9.0e-43 369 | R59281 y97g07.r1 Soares inf  
gb\_est4:H48290 - 445.00 907.78 8.5e-42 334 | H48290 y96g07.s1 Soares fet  
gb\_est11:AA232394 + 445.00 907.32 9.0e-42 349 | AA232394 zr27f02.r1 Stratage

seq\_name: gb\_est28:AI479776

seq\_documentation\_block: 720 bp. mRNA EST 14-APR-1999  
LOCUS AI479776  
DEFINITION tm69c04.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2163366 3'  
similar to TR:000264 000264 PUTATIVE PROGESTERONE BINDING PROTEIN.  
// mRNA sequence.

ACCESSION AI479776  
VERSION AI479776.1 GI:4372944  
KEYWORDS EST.  
SOURCE Homo sapiens

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 720)

AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index

JOURNAL Unpublished (1998)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3188831.

Contact: Robert Strausberg, Ph.D.

Seq primer: -400P from Gibco

High quality sequence stop: 469.

Location/Qualifiers

1..720

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2163366"

/clone\_lib="NCI\_CGAP\_Brn25"

/tissue\_type="anaplastic oligodendroglioma"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACATCTCAAGTGGAGCGGCATAGTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library is normalized, and was constructed by Bento

Soares and M.Fatima Bonaldo."

BASE COUNT 166 a 191 c 179 g 180 t 4 others

ORIGIN

alignment\_scores:

Quality: 902.00 Length: 211

Ratio: 4.602 Gaps: 3

Percent Similarity: 92.891 Percent Identity: 90.521

alignment\_block:

US-09-203-548-1 x AI479776/rev ..

Align seg 1/1 to reverse of: AI479776 from: 1 to: 720

13 ProSerAspLeuGluSerGlyGlyLeuHisGluIlePheThrSerPr 29



```

|||||
217 AGAACAAAAGATGAGAGTCCCGGAAAAATGATTAAAGCATTCAGTGA 168
|||||
201 SerileSer.*****yrPheAlaLysSerPheValThrValHis***v 217
|||||
167 AGTATATCTATTTTGTATTGTAATCATTTGTAACAGTCCACTCTG 118
|||||
217 alPheLysThr 220
|||||
117 TCITTAACA 107

```

seq\_name: gb\_est20:AA848529

seq\_documentation\_block: 534 bp mRNA EST 30-APR-1998  
 LOCUS AA848529 Normalized rat kidney, Bonto Soares Rattus sp. cDNA clone  
 DEFINITION RKIAC85 5' end similar to 25-Dx. mRNA sequence.

ACCESSION AA848529

VERSION AA848529.1 GI:2936069

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

Kerlavage,A.R. and Adams,M.D.

TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

JOURNAL Gene Index

COMMENT On Jan 19, 1998 this sequence version replaced gi:2286961.

Other\_ESTs: EST191288

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13 Reverse.

Location/Qualifiers

1..534

/organism="Rattus sp."

/db\_xref="ATCC (inhost):2007453"

/db\_xref="taxon:10118"

/clone="RKIAC85"

/note="Organ: kidney; Vector: pT73Pac; Site\_1: EcoRI;

Site\_2: NotI"

BASE COUNT 107 a 163 c 161 g 102 t 1 others

ORIGIN

alignment\_scores:

Quality: 729.00 Length: 157

Ratio: 4.926 Gaps: 0

Percent Similarity: 94.268 Percent Identity: 89.172

alignment\_block:

US-09-203-548-1 x AA848529

Align seg 1/1 to: AA848529 from: 1 to: 534

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17

|||||

63 ATGGCTCCGAGGATGTGGTGGCCGACTGGCCGCCACCCACGAGCTCGA 112

|||||

17 uSerGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34

|||||

113 GGGCGGGGGCTGCTCAAGAGATTTTCACGTCGCCTCTCAACCTGCTGC 162

|||||

34 euLeuGlyLeuGlyPheLeuLeuTyrlLysIleValArgGlyAspGln 50

|||||

```

163 TCCTGGCTCTGCATCTTCTCTCTACAAGATCGTTCCGGGGACCAG 212
|||||
51 ProAlaAlaSerGlyAspArgThrThr***ProProProLeuProAr 67
|||||
213 CCCGTCGACGTGGGACACGACGACGACGACGACGACGACGACGACG 262
|||||
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84
|||||
263 CCTCAAGCCGCGTGACTTCACCCCTGCCGACTAAGGGGATACGATGAG 312
|||||
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100
|||||
313 TCCAGGACCCGCGCATTTCTTATGGCCATCAACGCAAGGTGTCGACGTG 362
|||||
101 ThrLysGlyArgLysPheTyrglyProgluglyProtyrglyValPheAl 117
|||||
363 ACCAAGCCCGCAAGTTCTATGGCGGAGGGGCCATACGGGGTCTTTC 412
|||||
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134
|||||
413 TGGAAAGATGATCATCCAGNGCCCTTGGCCACATTTGCTGGACAAGAAG 462
|||||
134 laLeuLysAspGluTyrrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150
|||||
463 CACTGAAGATGAGTATGATGACCTTCTGACCTCACTCCTGCCGACGAG 512
|||||
151 GluThrLeuSerAspTrrpGlu 157
|||||
513 GGAGACCTGAATGACTGGGAC 533
|||||
seq_name: gb_est10:AA184910

```

seq\_documentation\_block:

LOCUS AA184910 535 bp mRNA EST 19-FEB-1997

DEFINITION mt62c03.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:634468 5'

similar to WP:K07E3.6 CE04722 TRANSLOCATING ATPASE ; mRNA

sequence.

ACCESSION AA184910

VERSION AA184910.1 GI:1768619

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 535)

AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilton,R. and

Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1281958.

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:386460

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 482.

FEATURES

source

1..535 Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:634468"

/clone\_lib="Soares 2NbMT"

/sex="male"

/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pMT7D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTACCAATCTGAAGTGGAGCGCGCGTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT731 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through two  
rounds of normalization, and was constructed by Bento  
Soares and M.Patima Bonaldo."  
BASE COUNT 103 a 171 c 159 g 102 t  
ORIGIN

alignment\_scores:  
Quality: 720.00 Length: 160  
Ratio: 4.675 Gaps: 0  
Percent Similarity: 96.250 Percent Identity: 92.500  
alignment\_block:  
US-09-203-548-1 x AA184910 ..  
Align seg 1/1 to: AA184910 from: 1 to: 535

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17  
58 ATGGCTGGCGAGGATGTGTGGCGACTGGCGC. GACCCGAGCGAGCTAGA 106  
17 userGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeu 34  
107 GGGCGGGCGGCTGTCACGAGATTTCACGCTCTCTCAACCTGCTCC 156  
34 euLeuGlyLeuCystIlePheLeuLeuTyrlLysIleValArgGlyAspGln 50  
157 TCCTGGGCTCTGCATCTTCCTGCTCTACAGATCGTTGCGGGGACGAG 206  
51 ProAlaAlaSerGlyAspArgThrThrThr\*\*\*ProProProLeuProAr 67  
207 CCCGTGCGAGTGGGACACGACGACGACGAA. CCACCCCGCGTGGCCCG 255  
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84  
256 CCTCAAGCGCGGCAACTTCACCCCTGCCAGCTGAGGCGCTTCGATGGCG 305  
84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100  
306 TCCAGGACCCGGCATCTCATGGCCATCAACGGCAAGGTGTCGACGTG 355  
101 ThrLysGlyArgLysPheTyrglyProGluGlyProTyrglyValPheAl 117  
356 ACCAAGGCGGCAAGTTCACGGGCTGAGGGGCACTATGGGCTCTTTCG 405  
117 aglyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134  
406 CGGAAGAGATGATCCAGGGGCTTGCACATTTGCTTGGCAAGAAG 455  
134 laLeuLysAspGluTyrrAspAspLeuSerAspLeuThrAlaAlaGlnGln 150  
456 CACTGAAGGATGAGTATGACGACCTTCTGACCTACCCCTGCACAGCAG 505  
151 GluThrLeuSerAspTrpGluSerGlnPhe 160  
506 AGTACCTCTGAGTGACTGGGACTCTCAGTTTC 535

seq\_name: gb\_est9:AA117455

seq\_documentation\_block:  
LOCUS AA117455 512 bp mRNA EST 15-NOV-1996  
DEFINITION mn28g11.r1 Beddington mouse embryonic region Mus musculus cDNA  
clone IMAGE:539300 5' similar to WP.K07E3.6 CE04722 TRANSLOCATING

ATPASE ;, mRNA sequence.  
ACCESSION AA117455  
VERSION AA117455.1 GI:1672468  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1393715.  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LInL : contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:326236

TITLE The WashU-HHMI Mouse EST Project

JOURNAL  
COMMENT

On Sep 12, 1996 this sequence version replaced gi:1393715.

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LInL : contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:326236

Seq primer: -40ml3 ET

High quality sequence stop: 466.

Location/Qualifiers

1. 512

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db\_xref="taxon:10090"

/clone="IMAGE:539300"

/clone\_lib="Beddington mouse embryonic region"

/sex="pooled"

/tissue\_type="embryo"

/dev\_stage="7.5dpc"

/lab\_host="DH12S"

/note="Organ: whole embryo; Vector: pCMV-SPORT; Site\_1:

SalI; Site\_2: NotI; Cloned unidirectionally. Primer:

Oligo dT. Gastrulating embryos were collected at 7.5dpc

from C57BL6 x DBA matings, excluding embryos that had

developed head folds and all extraembryonic tissues.

Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).

Referenced in Development 121, 2479-2489 (1995)"

134 a 128 c 132 g 117 t 1 others

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 709.00 Length: 160

Ratio: 4.727 Gaps: 1

Percent Similarity: 93.750 Percent Identity: 88.750

alignment\_block:

US-09-203-548-1 x AA117455 ..

Align seg 1/1 to: AA117455 from: 1 to: 512

62 ProProProLeuProArgLeuLysArgArgAspPheThrProAlaGluLe 78

|||||

6 CCACCCCGCTGCCCGCTCAAGCG.CGCACCTTCACCCCTGCCGAGCT 54

|||||

78 uArgArgPheAspGlyValGlnAspProArgIleLeuMetAlaIleAsnG 95

|||||

55 GAGGCGTTCGATGGCGTCCAGGACCCGCGCATCTCATGGCCATCAACG 104

|||||

95 lylLysValPheAspValThrLysGlyArgLysPheTyrglyProGluGly 111

|||||

105 GCAAGGTGTTCAGCTGACCAAGGCCCGCAAGTCTTACGGGCTGAGGGG 154

|||||







[5'pgactagttctatagtcgagcgccgctttttttttttttt3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 155 a 118 c 135 g 113 t 1 others  
ORIGIN

alignment\_scores:                   Length: 172  
                  Quality: 641.00                   Gaps: 1  
                  Ratio: 4.245  
Percent Similarity: 87.791   Percent Identity: 69.186

alignment\_block:

US-09-203-548-1 x AW018545

Align seg 1/1 to: AW018545 from: 1 to: 522

```

17  GluSerGlyLeuHisGluIlePheThrSerProLeuAsnLeuLe 33
18  ::::::::::::::::::::::::::::::::::::::::::::::
19  CAACATCTCGAATCTCTAGGAATTTTCAGTCGCCATCGACATCG 68
20  ::::::::::::::::::::::::::::::::::::::::::::::
21  uLeuLeuGlyLeuCysIlePheLeuLeuTyrlsIleValArgGlyAspG 50
22  ::::::::::::::::::::::::::::::::::::::::::::::
23  TTTGCTATGCTTGTGTGTTCTTACTTTACAAATCATCGCGGAGACA 118
24  ::::::::::::::::::::::::::::::::::::::::::::::
25  lnProAlaLysSerGlyAspArgThrThr***ProProProLeuPro 66
26  ::::::::::::::::::::::::::::::::::::::::::::::
27  AGCGTCGACACTATGCGC.....CGGTTGAGGAGCGCGCTGCC 156
28  ::::::::::::::::::::::::::::::::::::::::::::::
29  ArgLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspG 83
30  ::::::::::::::::::::::::::::::::::::::::::::::
31  AACTCAAGAAAGAGAGATTTTACTTTTACGACATCTGCAAGAGTACGATG 206
32  ::::::::::::::::::::::::::::::::::::::::::::::
33  yValGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspV 100
34  ::::::::::::::::::::::::::::::::::::::::::::::
35  ACTGAAACCCCAAGATCTCTGATGGTGTCTCAACGGGAAGATTTTGTATG 256
36  ::::::::::::::::::::::::::::::::::::::::::::::
37  alThrLysGlyArgLysPheTyrglyProGluGlyProTyrglyValPhe 116
38  ::::::::::::::::::::::::::::::::::::::::::::::
39  TTACAAGAGGGAAGAAATTTCTACGGTCCAGAGGCCCTTACGGAGTCTTC 306
40  ::::::::::::::::::::::::::::::::::::::::::::::
41  AlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysG 133
42  ::::::::::::::::::::::::::::::::::::::::::::::
43  CGCGGCAAGAGTCCGTCAGAGGCCCTGGCTACGTTCTGTTTGGAGAAGA 356
44  ::::::::::::::::::::::::::::::::::::::::::::::
45  uAlaLeuLysAspGluTyrlsAspAspLeuSerAspLeuThrAlaGlnG 150
46  ::::::::::::::::::::::::::::::::::::::::::::::
47  GGCTTGAAGACACTCATGATGATCTCTCGATCTCAATCCATGAGCAGC 406
48  ::::::::::::::::::::::::::::::::::::::::::::::
49  lngluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrlsHis 166
50  ::::::::::::::::::::::::::::::::::::::::::::::
51  AGGAGAGCCCTGAGCGAGTGGAGAGCCAGTTTCACAGAAAGTACGATTAC 456
52  ::::::::::::::::::::::::::::::::::::::::::::::
53  valGlyLysLeuLeuLysGluGluProThrValTyrlsSerAspG 183
54  ::::::::::::::::::::::::::::::::::::::::::::::
55  ATCGGGAAGCTTCTCAACCTGGAGAAGACCTACAGATACAGATGAG 506
56  ::::::::::::::::::::::::::::::::::::::::::::::
57  uGluGluProLysAsp 188
58  ::::::::::::::::::::::::::::::::::::::::::::::
59  NCGAGAAGTCAAGGAC 522

```

seq\_name: gb\_est28:AI477883

seq\_documentation\_block:

LOCUS AI477883 519 bp mRNA EST 09-MAR-1999  
DEFINITION fb57505.y1 zebrafish Washu MPIMG EST Danio rerio cDNA 5' similar to  
TR:000264 000264 PUTATIVE PROGESTERONE BINDING PROTEIN. ; mRNA  
sequence.

ACCESSION AI477883

VERSION AI477883.1 GI:4335494

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neoceratopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprininae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 519)

AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Willson, R.

TITLE Washu Zebrafish EST Project 1998

JOURNAL Unpublished (1998)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3186988.

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone Distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T3 ET from Amersham

High quality sequence stop: 493.

Location/Qualifiers

1..519

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone\_lib="zebrafish Washu MPIMG EST"

/sex="mixed"

/tissue\_type="26 somite embryos, adult livers, shield

stage embryos"

/lab\_host="XL1-blue MRF"

/notes="Vector: pSPORT1; Site.1: NotI; Site.2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dT)15 primer

[5'pgactagttctatagtcgagcgccgctttttttttttt3'];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPORT1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss), adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

155 a 116 c 135 g 113 t

BASE COUNT

ORIGIN

## alignment\_scores:

Quality: 640.00 Length: 169  
 Ratio: 4.267 Gaps: 1  
 Percent Similarity: 88.757 Percent Identity: 69.822

## alignment\_block:

US-09-203-548-1 x AI477883 ..

Align seg 1/1 to: AI477883 from: 1 to: 519

```

17 GluserGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLe 33
25 CAACCTTCGGAATCCTTCAGAAATTTTCAGTCGCCCACTCAACATCAG 74
33 uLeuLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspG 50
75 TTGCTATGCTTTGTTGTTCTTACTTTACAAATATCATCCCGGAGACA 124
50 lnProAlaAlaSerGlyAspArgThrThr***ProProProLeuPro 66
125 AGCCTGCAGATATGCG.....CCGTTGAGGAGCGCTGCC 162
67 ArgLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspG 83
163 AAACCTCAAGAAAGAGATTTTACTTTAGCAGATCTGCAAGAGTACGATG 212
83 yValGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspY 100
213 ACTGAAACCAACCAAGATCTCTGCTGCTCAACGGGAAAGATTGTATG 262
100 alThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPhe 116
263 TTACAAGAGGGAAGAAATCTACGCTCCAGAGGGCCCTTACGAGTCTTC 312
117 AlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysG 133
313 CGGGCAAGGATGCTCCAGAGGCTGCTGCTGCTGTTGGAGAAAGA 362
133 uAlaLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGlnG 150
363 GGCCTTGAAGACACATCATGATGATCTCCGATCTCAATGCATGCAGC 412
150 lnGluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrHisHis 166
413 AGGAGAGCTGAGCGAGTGGAGAGCCAGTTCACAGAGAGTACGATTAC 462
167 ValGlyLysLeuLeuLysGluGlyGluGluProThrValTyrSerAspG 183
463 ATCGGAAGCTTCTCAACCTGGAGAAACCTACAGAGTACACAGATCA 512
183 uGluGlu 185
513 CGAGGAA 519

```

seq\_name: gb\_est29:AI575080

## seq\_documentation\_block:

LOCUS AI575080 464 bp mRNA EST 05-APR-1999  
 DEFINITION UI-R-G0-ul-b-08-0-UI.s1 UI-R-G0 Rattus norvegicus cDNA clone  
 UI-R-G0-ul-b-08-0-UI 3', mRNA sequence.  
 ACCESSION AI575080  
 VERSION AI575080.1 GI:4559456  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 464)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477

## COMMENT

On May 18, 1998 this sequence version replaced gi:3138723.  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized ganglia library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-25, >AT-rich#Low\_complexity  
 Seq primer: M13 Forward.

## FEATURES

## source

1..464 Location/Qualifiers

/organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-G0-ul-b-08-0-UI"  
 /clone\_lib="UI-R-G0"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-G0 library is a normalized library constructed from a mixture of rat tissues (nodose ganglia, dorsal root ganglia, and trigeminal ganglia). The tag is a string of 6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996."  
 BASE COUNT 108 a 122 c 98 g 136 t  
 ORIGIN

## alignment\_scores:

Quality: 639.00 Length: 143  
 Ratio: 4.769 Gaps: 1  
 Percent Similarity: 93.706 Percent Identity: 86.014

## alignment\_block:

US-09-203-548-1 x AI575080/rev ..

Align seg 1/1 to reverse of: AI575080 from: 1 to: 464

```

79 ArgArgPheAspGlyValGlnAspProArgIleLeuMetAlaIleAsnG 95
462 AGCGATACGATGGAGTCCAGGACCGCGCATCTTATGGCCATCAACGG 413
95 yLysValPheAspValThrLysGlyArgLysPheTyrGlyProGluGlyP 112
412 CAAGGTGTTTCGACGTGACCAAGCGCGCAAGTCTATGGCGCGAGGGGC 363
112 rGlyGlyValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPhe 128
362 CATACGGGTCTTCTGCTGGAAGAGATGTCATCCAGGGGCCCTGCCACAT 313
129 CysLeuAspLysGluAlaLeuLysAspGluTyrAspAspLeuSerAsp 145
312 TGCCTGGCAAGAGAGCACTGAAGGATGATGATGACCTTCTGACCT 263
145 uThrAlaAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPheThrP 162
262 CACTCCTGCCAGCAGGAGACCCCTGAATGACTGGGACTCTCAGTTACCT 213
162 heLysTyrHisHisValGlyLysLeuLysGluGlyGluGluProThr 178
212 TCAGTACCATCAGTGGGAAACTGCTGAAGAGAGGGAGGCGGACT 163

```

179 valTyrSeraspGluGluGluProLysaspGluSerSerArgLyAsn.V 195  
:  
162 GTCTACTCGATGATGAAGAACCAAAAGATCAGCGTGCTCGAAGAGTA 113  
:  
195 allYsAlaPheSerglySerileSer\*\*\*\*\*TyrPheAlaylSerPhe 211  
:  
112 CTGAAGCAGTCAGTGAGCATATCTATTTTGTATTTTGTCAAAATCAVTT 63

```

seq_documentation_block:
LOCUS       N52291               458 bp    mRNA                    EST
DEFINITION  yv48a12.s1 Soares fetal liver spleen 1NF5S Homo sapiens cDNA clone
IMAGE:245950 3', mRNA sequence.

```

REFERENCE  
AUTHORS

1 (bases 1 to 458)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R., and Marra, M.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CONSTRUCTED BY BENCO SOARES AND M.FACILIND BONADIO:			
BASE COUNT	102 a	126 c	114 a 114 t 2 others

alignment scores:

Quality:	636.00	Length:	155
Ratio:	4.417	Gaps:	5
Percent Similarity:	92.903	Percent Identity:	91.613

```
alignment_block:
```

US-09-203-548-1 x N52291/rev

Align seq 1/1 to reverse of: N52291 from: 1 to: 458

68 LeuLYsArgArGAspPheThr . ProMaGIuLeuArgArGpHeAspGlyV 84  
457 CTCACGGCGGGACTTCACCCCGCCGAGCTGGCGGCTTCGACGGG 408  
84 aGIlnAspPro..ArgIleLeuMetAlaIleAsnGlyLYsValPHeAspV 100  
407 TCCAGGACCCCGGCCCATCTCATGGCCATCAACGCAAGGTGTTCTGATG 358  
100 aUthrLYsGlyArg .LYsPheTyrGlyProGluGlyProTyrGlyValPh 116  
357 TGACCAAAAGCCGCANAATTCACCGGCGNAGGGGCCGTATGGGGTCTT 308  
116 eAlaGlyArgAspAlaSerArGlyLeuAlaThrPheCysLeuAspLYsG 133  
307 TGTGTGAAGAGATGCAATCCAG .GGCCTCCACATTTTGCCTGGATAAG 259  
133 LuAlaLeuLYsAspGluTyrAspAspLeuSerAspLeuThrAlaAlaGln 149  
258 AGCACTGAAGATGATGATACGATGACCTTCTGACCTCACTGCTGCCAG 209  
150 GlnGluThrLeuSerAspTrpGluSerGlnPheThrPheLYsTyrHisH 166  
208 CAGGAGACTCTGAGTGACTGGAGTGCTCAGTTCACCTTTCAAGTATCAT 159  
166 sValGlyLYsLeuLeuLYsGluGlyGluGluProThrValTyrSerAsp 183  
158 CGTGGCAAACTGCTCAAGAGGGGGAGGAGGCCACTGTGTACTCAGATG 109  
183 luclGluProLYsAspGluSerSerArGlyAsn .ValLYsAlaPheSe 199  
108 AGGAAGAACCAAAAGATGAGATGGCCGGGAAAATGATTAAAGCAPTTCAG 59  
199 rGlySerIleSer .\*\*\*\*\*TyrPHeAlaLYsSerPHeValThrValHis 215  
58 TGAAGTATATCTATTTTTGTATTGTAATAATCATTTTGTACAGTCCAC 9  
216 \*\*\*val 217  
8 TCTGTC 3

## COMMENT

On Jun 5, 1998 this sequence version replaced gi:3189370.  
 Other ESTs: fc43g04.xl  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watson.wustl.edu  
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 Resourcezentrum Primatendatenbank, Berlin, Germany (web address:  
 www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 499.

## FEATURES

source

Location/Qualifiers

1. .563  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone\_lib="Zebrafish WashU MPIMG EST"  
 /sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield  
 stage embryos"  
 /lab\_host="XL1-blue MRF"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
 strand cDNA was primed with a Not I - oligo(dT)15 primer  
 [5'-pgactagttctagatcgagcgccgctttttttttttttt3'];  
 double-stranded cDNA was ligated to Sal I adaptors (BRL),  
 digested with Not I and cloned into the Not I and Sal I  
 sites of the pSPORT1 vector (BRL). Library was constructed  
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
 analysis were selected following oligonucleotide  
 hybridization fingerprinting of arrayed clones from  
 zebrafish late somitogenesis (26 ss), adult liver or  
 embryonic shield stage (5.6 h) libraries. Fingerprint  
 data were used to computationally cluster cDNAs, and a  
 single cDNA from each cluster was chosen for sequencing.  
 In some cases multiple members of the same cluster were  
 sequenced to assess clustering parameters or single clones  
 were sequenced additional times to assess quality  
 control."

BASE COUNT 171 a 126 c 148 g 117 t 1 others  
 ORIGIN

## alignment\_scores:

Quality: 636.00 Length: 173  
 Ratio: 4.212 Gaps: 1  
 Percent Similarity: 87.283 Percent Identity: 68.208

## alignment\_block:

US-09-203-548-1 x AI794380 ..

Align seg 1/1 to: AI794380 from: 1 to: 563

17 GluSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAspLeuLe 33  
 :::  
 22 CAACCTCTCGAATCCTTCAGGAATTTTCACGTCGCCACCTGAACATCGG 71  
 :::  
 33 uLeuLeuGlyLeuCysIlePheLeuLeuTyrlsIleValArgGlyAspG 50  
 :::  
 72 TTGGCTATGCTTTGTGTTCCTACITTTACAAATCATCTCCGGGAGACA 121  
 :::  
 50 lnProAlaIaSerGlyAspArgThrThr\*\*\*ProProProLeuPro 66  
 :::  
 122 AGCCTGCAGACATATGCC.....CCGGTTGAGGACCGCTGCC 159  
 :::  
 67 ArgLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspG1 83  
 :::

160 AAACCTCAAGAAAGAGATTTTACTTTAGCAGATCTCCRAAGAGTACGATGG 209  
 83 yValGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspV 100  
 :::  
 210 ACTGAAACACCCCAAGATCTCTGATGCTGTCAACGGGAAAGTATTGTATG 259  
 100 alThrLysGlyArgLysPheTyrglyProGluGlyProTyrglyValPhe 116  
 :::  
 260 TTACAGAGGGGAAATTTCTACGGTCCAGAGGCCCTTACGGAGTCTTC 309  
 117 AlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysG1 133  
 :::  
 310 GCGGGCAAGGATGCGTCAGAGAGCCCTGGCTCTCTGTTTGGAGAAAGA 359  
 133 uAlaLeuLysAspGluTyArgAspLeuSerAspLeuThrAlaAlaGlnG 150  
 :::  
 360 GGCCTTGAAGACACTCATGATGATCTCTCCGATCTCAATGCCATGCAGC 409  
 150 lnGluThrLeuSerAspTrpGluSerGlnPheThrPheLysTyrlsHis 166  
 :::  
 410 AGGAGAGCCTGAGCGAGTGGGAGACCTTCACACAGAGTACGATTAC 459  
 167 ValGlyLysLeuLeuLysGluGlyGluGluProThrValTySerAspG1 183  
 :::  
 460 ATCGGAAGCTTCTCANACCTGGAGAGAACCTACAGACTACACAGATGA 509  
 183 uGluGluProLysAspGlu 189  
 :::  
 510 CGAGGAAGTCAAGGACAAA 528

seq\_name: gb\_est12:AA288129

## seq\_documentation\_block:

LOCUS AA288129 508 bp mRNA EST 11-APR-1997  
 DEFINITION vbl2e11.r1 Soares mouse NML Mus musculus CDNA clone IMAGE:748748 5'  
 similar to TR:8247050 E247050 CHROMOSOME XVI READING FRAME ORF  
 YPL170W. ;, mRNA sequence.

ACCESSION AA288129  
 VERSION AA288129.1 GI:1934364  
 KEYWORDS EST.  
 SOURCE house musculus  
 ORGANISM Mus musculus

## REFERENCE

AUTHORS

1 (bases 1 to 508)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,I., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

## TITLE

The WashU-HMI Mouse EST Project

## JOURNAL

Unpublished (1996)

## COMMENT

On Sep 12, 1996 this sequence version replaced gi:1397407.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of MedicineP  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:457732

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 465.

## FEATURES

source

Location/Qualifiers

1. .508  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:748748"  
 /clone\_lib="Soares mouse NML"  
 /tissue\_type="Liver"  
 /lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATGTCAGTGGAGCGGCCCGAATCTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 97 a 163 c 154 g 94 t

ORIGIN

alignment\_scores: Quality: 630.00 Length: 149  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 96.644 Percent Identity: 94.631

alignment\_block:

US-09-203-548-1 x AA288129

Align seg 1/1 to: AA288129 from: 1 to: 508

1 MetAlaAlaGluAspValValAlaThrGlyAlaAspProSerAspLeuG1 17  
|||||  
64 ATGGCTCCGAGGATGTTGGCGACTGGCCGCCACCGAGCAGCTAGA 113  
|||||  
17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34  
|||||  
114 GGGCGCGGGCTGCTGCACGAGATTTTCAGTCTCTCTCAACCTGCTCC 163  
|||||  
34 euLeuGlyLeuCysIlePheLeuLeuTyrlsyleValArgGlyAspGln 50  
|||||  
164 TCGTGGCGCTCTGCATCTTCTCTCTACAGATCGTTCCGGGGACCAG 213  
|||||  
51 ProAlaAlaSerGlyAspArgThrThrThr\*\*\*ProProProLeuProAr 67  
|||||  
214 CCGGTGCCAGTGGGACACGACGACGACGAA .CCACCCCGCTGCCCG 262  
|||||  
67 gLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84  
|||||  
263 CCTCAAGCG .CGGACTTCACCCCTGCCGAGCTGAGCGTTTCGATGGCG 311  
|||||  
84 aGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100  
|||||  
312 TCAGGACCGCGCATCTCATGGCCATCAACGCAAGGTGTTCCAGCGT 361  
|||||  
101 ThrLysGlyArgLysPheTyrlsyleProGluGlyProTyrlsyleValPheAl 117  
|||||  
362 ACCAAGGCGCGAAGTCTACGGGCTGAGGGCA .TATGGGTCTTTGC 410  
|||||  
117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134  
|||||  
411 CGGAAGATGCATCCAGGGCCCTTGC .ACATTTGCCTGCACAAGAAG 459  
|||||  
134 laLeuLysAspGluTyrlsyleAspLeuSerAspLeuThrAlaAlaGln 149  
|||||  
460 CACTGAAGGATGAGTATGACGACCTTTCTGACCTCACCCCTGCACAG 506  
|||||

seq\_name: gb\_est27:A1453643

seq\_documentation\_block:

LOCUS A1453643 525 bp mRNA EST 13-APR-1999  
DEFINITION t145h05.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone IMAGE:2144505 3' similar to TR:000264 O00264 PUTATIVE PROGESTERONE BINDING PROTEIN. ; mRNA sequence.

ACCESSION A1453643  
VERSION A1453643.1 GI:4283551  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE

1 (bases 1 to 525)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

# JOURNAL

Unpublished (1997)

On Mar 10, 1998 this sequence version replaced gi:2948746.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Insert Length: 696 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 399.

# FEATURES

source

1..525  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2144505"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 129 c 103 g 156 t.

# alignment\_scores:

Quality: 625.00 Length: 142  
Ratio: 4.808 Gaps: 2  
Percent Similarity: 91.549 Percent Identity: 88.732

# alignment\_block:

US-09-203-548-1 x A1453643/rev

Align seg 1/1 to reverse of: A1453643 from: 1 to: 525

81 PheAspGlyValGlnAspProArgIleLeuMetAlaIleAsnGlyLysVa 97  
|||||  
525 TTGTACCGCTCCAGGACCGCGCAATATCATGGCCATCACGGCAAGG 476  
|||||  
97 lPheAspValThrLysGlyArgLysPheTyrlsyleProGluGlyProTyrg 114  
|||||  
475 TGTTTCGATGTACCAAGGCGCAAAATTTACGGGCGCGAGGGCCGTATG 426  
|||||  
114 lyValPheAlaGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeu 130  
|||||  
425 GGGTCTTTGCTGGGAAGAGATGCATCCAGGGCCCTTGCCACATTTTGCCTG 376  
|||||  
131 AspLysGluAlaLeuLysAspGluTyrlsyleAspLeuSerAspLeuThrAl 147  
|||||  
375 GATAAGGAACACCTGAGGATGAGTACGATGACCTTCTGACCTCACTGC 326  
|||||  
147 aAlaGlnGlnGluThrLeuSerAspTrpGluSerGlnPheThrPheLysT 164  
|||||  
325 TGCCACGACGAGACTCTGAGTACTGGAGTCTCAGTTCACTTTCAAGT 276  
|||||  
164 yrHisHsValGlyLysLeuLeuLysGluGlyGluProThrValTyrlr 180  
|||||

275 ATCATCAGTGGGCAAACTGCTGAAGGAGGGGAGGAGCCCACTGTGTAC 226  
 181 SerAspGluGluProLysAspGluSerArgLysAsn.ValLysA 197  
 225 TCAGATGAGGAAGAACCAAGATGAGATGCTCCCGGAAATGATTAAG 176  
 197 laPheSerGlySerIleSer.\*\*\*\*\*TyrPheAlaLysSerPheValTh 213  
 175 CATTCAGTGGGAAGTATATCTATTGTTATTTTCAAAATCATTTGTAAC 126  
 213 rValHis\*\*\*ValPheLysThr 220  
 125 AGTCCTCTCTCTTTAAACA 104

seq\_name: gb\_est14:AA428520

seq\_documentation\_block: 531 bp mRNA EST 16-OCT-1997  
 LOCUS AA428520 zw47c07.r1 Soares.total\_fetus\_Nb2HF8\_9w Homo sapiens cDNA clone  
 DEFINITION IMAGE:773196 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING ATPASE  
 ; mRNA sequence.

ACCESSION AA428520  
 VERSION AA428520.1 GI:2112535  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 531)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
 Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wyllie,T., Waterston,K. and Wilson,R.  
 WashU-Merck EST Project 1997  
 Unpublished (1997)

TITLE On Sep 12, 1996 this sequence version replaced gi:1394912.  
 JOURNAL Contact: Willson RK  
 COMMENT Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28ml3 rev2 Et from Amersham  
 High quality sequence stop: 475.

#### FEATURES

Location/Qualifiers  
 1..531  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:773196"  
 /clone\_lib="Soares.total\_fetus\_Nb2HF8\_9w"  
 /dev\_stage="8-9 weeks"  
 /lab\_host="DH10B"  
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from pooled 8-9 week  
 (total) fetus material with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGGAGCGGCTTAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo. "  
 BASE COUNT 99 a 172 c 162 g 95 t 3 others  
 ORIGIN

alignment\_scores:  
 Quality: 622.50 Length: 148  
 Ratio: 4.544 Gaps: 1  
 Percent Similarity: 92.568 Percent Identity: 90.541

alignment\_block:  
 US-09-203-548-1 x AA428520 ..  
 Align seg 1/1 to: AA428520 from: 1 to: 531  
 1 MetAlaAlaGluAspValAlaThrGlyAlaAspProSerAspLeuGl 17  
 92 ATGGCTGCCGAGGATGTGTGGCGACTGGCGC.GACCCAAGCGATCTGA 140  
 17 uSerGlyGlyLeuLeuHisGluIlePheThrSerProLeuAsnLeuLeuL 34  
 141 GAGCGCGGGCTGCTGCATGAGATTTTCAGTCGCCGCTCAACCTGCTGC 190  
 34 euLeuGlyLeuCysIlePheLeuLeuTyrLysIleValArgGlyAspGln 50  
 191 TCGTTGGGCTCTGCATCTTCTGCTCTACAGATCTGCGGGGGACCAG 240  
 51 ProAlaAlaSerGlyAspArgThrThrThr\*\*\*ProProProLeuProAr 67  
 241 CGGCGGCGAGCG...GANCACGACGACGAGCGGCCCTCTGCCCGC 287  
 67 gluLeuLysArgArgAspPheThrProAlaGluLeuArgArgPheAspGlyV 84  
 288 CCTCAAGCGCGCGANCTTCACCCCGCGAGCTGCGGCGCTTCGACGCG 337  
 84 alGlnAspProArgIleLeuMetAlaIleAsnGlyLysValPheAspVal 100  
 338 TCCAGGACCGCGCATACTCATGGCCATCAAGCGCAAGGTGTTCGATGTG 387  
 101 ThrLysGlyArgLysPheTyrGlyProGluGlyProTyrGlyValPheAl 117  
 388 ACCAAGGCCCGCAATTTACGGGCGCCGAGCGGTA.NTGGGGGTCTTTGC 436  
 117 aGlyArgAspAlaSerArgGlyLeuAlaThrPheCysLeuAspLysGluA 134  
 437 TGGAAAGAGATGCATCCAGGGCGCTTGCACAAATTTTCTGGATAAGAAG 486  
 134 laLeuLysAspGluTyrAspAspLeuSerAspLeuThrAlaAla 148  
 487 CACTGAAGGATGAGTACGATGACCTTTCTGACCTCACTGCTGCC 530



\*\*\*\*\* STN Columbus \*\*\*\*\*

FILE 'HOME' ENTERED AT 18:26:24 ON 16 JAN 1998

=> index bioscience

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	
	ENTRY SESSION
FULL ESTIMATED COST	0.15 0.15

INDEX 'AGRICOLA, AIDSLINE, ANABSTR, AQUASCI, BIOBUSINESS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CANCERLIT, CAPLUS, CEABA, CEN, CIN, CJACS, CJELSEVIER, CONFSCI, CROPB, CROPU, DDFB, DDFU, DGENE, DISSABS, DRUGB, DRUGLAUNCH, DRUGNL, DRUGU, EMBAL, EMBASE, ...' ENTERED AT 18:26:36 ON 16 JAN 1998

50 FILES IN THE FILE LIST IN STNINDEX

Enter SET DETAIL ON to see search term postings or to view search error messages that display as 0\* with SET DETAIL OFF.

=> s cystar? and (cytokine? or steroid?)

22 FILES SEARCHED...  
36 FILES SEARCHED...

0 FILES HAVE ONE OR MORE ANSWERS, 50 FILES SEARCHED IN STNINDEX

L1 QUE CYSTAR? AND (CYTOKINE? OR STEROID?)

=> s steroid? and cytokine?

7 FILE AGRICOLA  
32 FILE AIDSLINE  
1 FILE AQUASCI  
12 FILE BIOBUSINESS  
638 FILE BIOSIS  
7 FILE BIOTECHABS  
7 FILE BIOTECHDS  
47 FILE CABA  
687 FILE CANCERLIT  
941 FILE CAPLUS  
2 FILE CEABA  
3 FILE CEN  
3 FILE CIN  
29 FILE CJACS  
5 FILE CONFSCI  
1 FILE DDFB  
107 FILE DDFU  
25 FILE DISSABS  
1 FILE DRUGB  
19 FILE DRUGNL  
223 FILE DRUGU  
18 FILE EMBAL  
932 FILE EMBASE  
1 FILE GENBANK  
1 FILE HEALSAFE  
18 FILE IFIPAT  
2016 FILE JICST-EPLUS  
2 FILE JPNEWS  
167 FILE LIFESCI  
1025 FILE MEDLINE  
1 FILE NIOSHTIC  
7 FILE NTIS  
2 FILE PHAR  
40 FILE PHIN

112 FILE PROMT  
795 FILE SCISEARCH  
309 FILE TOXLINE  
240 FILE TOXLIT  
619 FILE USPATFULL  
44 FILE WPIDS  
44 FILE WPINDEX

41 FILES HAVE ONE OR MORE ANSWERS, 50 FILES SEARCHED IN STNINDEX

L2 QUE STEROID? AND CYTOKINE?

=> s (steroid? or progesterone?)(5a) (membrane bind? protein?)

1 FILE AGRICOLA  
4 FILE BIOSIS  
6 FILES SEARCHED...  
10 FILES SEARCHED...  
3 FILE CAPLUS  
21 FILES SEARCHED...  
3 FILE EMBASE  
29 FILES SEARCHED...  
5 FILE GENBANK  
2 FILE LIFESCI  
36 FILES SEARCHED...  
3 FILE MEDLINE  
37 FILES SEARCHED...  
4 FILE SCISEARCH  
46 FILES SEARCHED...  
3 FILE TOXLIT

9 FILES HAVE ONE OR MORE ANSWERS, 50 FILES SEARCHED IN STNINDEX

L3 QUE (STEROID? OR PROGESTERONE?)(5A) (MEMBRANE BIND? PROTEIN?)

=> file biosis medline caplus scisearch genbank agricola

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	
	ENTRY SESSION
FULL ESTIMATED COST	9.90 10.05

FILE 'BIOSIS' ENTERED AT 18:39:47 ON 16 JAN 1998  
COPYRIGHT (C) 1998 BIOSIS(R)

FILE 'MEDLINE' ENTERED AT 18:39:47 ON 16 JAN 1998

FILE 'CAPLUS' ENTERED AT 18:39:47 ON 16 JAN 1998  
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
COPYRIGHT (C) 1998 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'SCISEARCH' ENTERED AT 18:39:47 ON 16 JAN 1998  
COPYRIGHT (C) 1998 Institute for Scientific Information (ISI) (R)

FILE 'GENBANK' ENTERED AT 18:39:47 ON 16 JAN 1998

FILE 'AGRICOLA' ENTERED AT 18:39:47 ON 16 JAN 1998

=> s l3

2 FILES SEARCHED...  
L4 20 L3

=> dup rem l4



DUPLICATE IS NOT AVAILABLE IN 'GENBANK'.  
ANSWERS FROM THESE FILES WILL BE CONSIDERED  
UNIQUE  
PROCESSING COMPLETED FOR L4  
L5 12 DUP REM L4 (8 DUPLICATES REMOVED)

=> d 1-12

L5 ANSWER 1 OF 12 SCISEARCH COPYRIGHT 1998 ISI (R)  
AN 97:201327 SCISEARCH  
GA The Genuine Article (R) Number: WL530  
TI Purification and sequencing of a **progesterone  
membrane binding protein**.  
AU Falkenstein E (Reprint); Meyer C; Wehling M  
CS UNIV HEIDELBERG, FAC CLIN MED, INST CLIN  
PHARMACOL, D-68167  
MANNHEIM, GERMANY; UNIV MUNICH, INST CLIN  
PHARMACOL, D-80336 MUNICH,  
GERMANY  
CYA GERMANY  
SO FASEB JOURNAL, (28 FEB 1997) Vol. 11, No. 3, pp. 1875-  
1875.

Publisher: FEDERATION AMER SOC EXP BIOL, 9650  
ROCKVILLE PIKE,  
BETHESDA, MD 20814-3998.  
ISSN: 0892-6638.

DT Conference; Journal  
FS LIFE  
LA English  
REC Reference Count: 0

L5 ANSWER 2 OF 12 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 97:185563 BIOSIS  
DN 99484766

TI Purification and sequencing of a **progesterone  
membrane binding protein**.  
AU Falkenstein E; Meyer C; Wehling M  
CS Inst. Clin. Pharmacol., Fac. Clin. Med., Mannheim, Univ.  
Heidelberg,  
68167 Mannheim, Germany  
SO Annual Meeting of the Professional Research Scientists on  
Experimental Biology 97, New Orleans, Louisiana, USA, April 6-  
9,  
1997. FASEB Journal 11 (3). 1997. A323. ISSN: 0892-6638  
DT Conference  
LA English

L5 ANSWER 3 OF 12 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 97:538761 BIOSIS  
DN 99837964

TI Characterization, purification and sequencing of a  
**progesterone membrane binding  
protein**.  
AU Wehling M; Meyer C; Gerdes D; Falkenstein E  
CS Inst. Clin. Pharmacol., Fac. Clin. Med. at Mannheim, Univ.  
Heidelberg, Heidelberg, Germany  
SO 2nd Congress of the European Association for Clinical  
Pharmacology  
and Therapeutics, Berlin, Germany, September 17-20, 1997.  
European  
Journal of Clinical Pharmacology 52 (SUPPL.). 1997. A79.  
ISSN:  
0031-6970  
DT Conference  
LA English

L5 ANSWER 4 OF 12 SCISEARCH COPYRIGHT 1998 ISI (R)  
AN 97:339100 SCISEARCH  
GA The Genuine Article (R) Number: WV420

TI Characterization, purification and sequencing of a  
**progesterone membrane binding  
protein**.

AU Falkenstein E (Reprint); Meyer C; Wehling M  
CS UNIV HEIDELBERG, INST CLIN PHARMACOL, FAC  
CLIN MED MANNHEIM, D-68167  
MANNHEIM, GERMANY; UNIV MUNICH, KLINIKUM  
INNENSTADT, MED KLIN, INST  
CLIN PHARMACOL, D-80336 MUNICH, GERMANY  
CYA GERMANY  
SO NAUNYN-SCHMIEDEBERGS ARCHIVES OF  
PHARMACOLOGY, (1 MAY 1997) Vol.  
355, No. 4, Supp. [S], pp. 56-56.  
Publisher: SPRINGER VERLAG, 175 FIFTH AVE, NEW  
YORK, NY 10010.  
ISSN: 0028-1298.

DT Conference; Journal  
FS LIFE  
LA English  
REC Reference Count: 0

L5 ANSWER 5 OF 12 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 1  
AN 97:42629 BIOSIS  
DN 99334617

TI Full-length cDNA sequence of a **progesterone  
membrane-binding protein** from porcine  
vascular smooth muscle cells.  
AU Falkenstein E; Meyer C; Eisen C; Scriba P C; Wehling M  
CS Division Clinical Pharmacol., Med. Klinik, Klinikum Innenstadt,  
Univ.  
Munich, Ziemssenstr. 1, 80336 Munich, Germany  
SO Biochemical and Biophysical Research Communications 229  
(1). 1996.  
86-89. ISSN: 0006-291X  
LA English

L5 ANSWER 6 OF 12 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 2  
AN 93:231646 BIOSIS  
DN BA95:122821  
TI A SPECIFIC MEMBRANE BINDING PROTEIN FOR  
**PROGESTERONE** IN RAT BRAIN SEX DIFFERENCES AND  
INDUCTION BY  
ESTROGEN.  
AU TISCHKAU S A; RAMIREZ V D  
CS DEP. PHYSIOL. BIOPHYSICS, UNIV. ILLINOIS URBANA-  
CHAMPAIGN, 524  
BURRILL HALL, 407 SOUTH GOODWIN AVE., URBANA,  
IL 61801, USA.  
SO PROC NATL ACAD SCI U S A 90 (4). 1993. 1285-1289.  
CODEN: PNASA6  
ISSN: 0027-8424  
LA English

L5 ANSWER 7 OF 12 MEDLINE DUPLICATE  
3  
AN 89054044 MEDLINE  
DN 89054044

TI Insulin and epidermal growth factor stimulate phosphorylation of  
a  
170-kDa protein in intact hepatocytes immunologically related to  
lipocortin 1.  
AU Karasik A; Pepinsky R B; Kahn C R  
CS Research Division, Joslin Diabetes Center, Boston,  
Massachusetts  
02215.  
NC DK33201 (NIDDK)  
SO JOURNAL OF BIOLOGICAL CHEMISTRY, (1988 Dec 5)  
263 (34) 18558-62.

Journal code: HIV. ISSN: 0021-9258.  
CY United States  
DT Journal; Article; (JOURNAL ARTICLE)  
LA English  
FS Priority Journals; Cancer Journals  
EM 198903

L5 ANSWER 8 OF 12 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): AA705751 GenBank (R)  
GenBank ACC. NO. (GBN): AA705751  
SEQUENCE LENGTH (SQL): 390  
MOLECULE TYPE (CI): mRNA; linear  
DIVISION CODE (CI): Expressed sequence tag  
DATE (DATE): 24 Dec 1997  
DEFINITION (DEF): zf41c07.s1 Soares fetal heart NbHH19W  
Homo

sapiens cDNA clone 379500 3' similar to TR:Q95250  
Q95250 STEROID MEMBRANE

**BINDING PROTEIN. ;**

KEYWORDS (ST): EST

SOURCE: human.

ORGANISM (ORGN): Homo sapiens  
Eukaryotae; Metazoa; Chordata; Vertebrata;  
Mammalia; Eutheria; Primates; Catarrhini;  
Hominidae; Homo

NUCLEIC ACID COUNT (NA): 79 a 112 c 130 g 67 t 2  
others

COMMENT:

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further  
information.

Possible reversed clone: similarity on wrong strand

Possible reversed clone: polyT not found

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 377.

REFERENCE: 1 (bases 1 to 390)

AUTHOR (AU): Hillier,L.; Allen,M.; Bowles,L.;  
Dubuque,T.;

Geisel,G.; Jost,S.; Krizman,D.; Kucaba,T.;  
Lacy,M.; Le,N.; Lennon,G.; Marra,M.; Martin,J.;  
Moore,B.; Schellenberg,K.; Steptoe,M.; Tan,F.;  
Theising,B.; White,Y.; Wylie,T.; Waterston,R.;  
Wilson,R.

TITLE (TI): WashU-NCI human EST Project

JOURNAL (SO): Unpublished (1997)

L5 ANSWER 11 OF 12 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): ATFCA2 GenBank (R)  
GenBank ACC. NO. (GBN): Z97337  
CAS REGISTRY NO. (RN): 194251-50-8  
SEQUENCE LENGTH (SQL): 202861  
MOLECULE TYPE (CI): DNA; linear  
DIVISION CODE (CI): Plants, fungi, algae  
DATE (DATE): 4 Jul 1997  
DEFINITION (DEF): Arabidopsis thaliana DNA chromosome 4,  
ESSA I

contig fragment No. 2.

SOURCE: thale cress.

ORGANISM (ORGN): Arabidopsis thaliana  
Eukaryotae; mitochondrial eukaryotes;  
Viridiplantae; Charophyta/Embryophyta group;

Embryophyta; vascular plants; seed plants;  
Magnoliophyta; Magnoliopsida; Capparales;  
Brassicaceae; Arabidopsis

NUCLEIC ACID COUNT (NA): 66100 a 35874 c 36335 g  
64552 t

COMMENT:

this fragment has an overlap with ATFCA1 at the 5' end and an  
overlap  
with ATFCA3 at the 3' end.

REFERENCE: 1 (bases 1 to 202861)

AUTHOR (AU): Bevan,M.; Stiekema,W.; Murphy,G.;  
Wambutt,R.;

Pohl,T.; Terry,N.; Kreis,M.; Kavanagh,T.;  
Entian,K.D.; Rieger,M.; James,R.;  
Puigdomenech,P.; Hatzopoulos,P.; Obermaier,B.;  
Duesterhoft,A.; Jones,J.; Palme,K.; Ansoorge,W.;  
Delseny,M.; Bancroft,I.; Mewes,H.W.;  
Schueller,C.; Chalwatzis,N.

JOURNAL (SO): Unpublished

REFERENCE: 2 (bases 1 to 202861)

AUTHOR (AU): EU Arabidopsis sequencing project; ESSA.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (19-JUN-1997) MIPS, at the  
Max-Planck-Institut fuer Biochemie, Am  
Klopferspitz 18a, D-82152 Martinsried, FRG,  
Project Coordinator: Mike Bevan, Molecular  
Genetics Department, Cambridge Laboratory, John  
Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk

L5 ANSWER 12 OF 12 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): SSSTERMBP GenBank (R)

GenBank ACC. NO. (GBN): X99714

CAS REGISTRY NO. (RN): 182911-41-7

SEQUENCE LENGTH (SQL): 1893

MOLECULE TYPE (CI): RNA; linear

DIVISION CODE (CI): Other mammals

DATE (DATE): 11 Jul 1997

DEFINITION (DEF): S.scrofa mRNA for steroid  
membrane binding  
protein.

KEYWORDS (ST): steroid membrane  
binding protein

SOURCE: pig.

ORGANISM (ORGN): Sus scrofa  
Eukaryotae; mitochondrial eukaryotes; Metazoa;  
Chordata; Vertebrata; Mammalia; Eutheria;  
Artiodactyla; Suiformes; Suina; Suidae; Sus

NUCLEIC ACID COUNT (NA): 509 a 447 c 461 g 476 t

COMMENT:

Reference: Meyer, C.; J.Biol.Chem. 239; 726-731; 1996 (N-  
terminal  
protein-sequence).

REFERENCE: 1 (bases 1 to 1893)

AUTHOR (AU): Falkenstein,E.; Meyer,C.; Eisen,C.;  
Scriba,P.C.;

Wehling,M.

TITLE (TI): Full-length cDNA sequence of a  
progesterone membrane-  
binding protein from porcine  
vascular smooth muscle cells

JOURNAL (SO): Biochem. Biophys. Res. Commun., 229 (1),  
86-89

(1996)

OTHER SOURCE (OS): CA 126:42820

REFERENCE: 2 (bases 1 to 1893)

AUTHOR (AU): Falkenstein,E.

TITLE (TI): Direct Submission  
JOURNAL (SO): Submitted (01-AUG-1996) E. Falkenstein,  
University of Heidelberg, Institute of Clinical  
Pharmacology, Faculty of Clinical Medicine at  
Mannheim, Theodor-Kutzer-Ufer, 68135 Mannheim,  
FRG

=> e selmin o/au

E1 4 SELMIKAT I/AU  
E2 1 SELMIN D W/AU  
E3 46 -> SELMIN O/AU  
E4 8 SELMIN ORNELLA/AU  
E5 6 SELMIN V/AU  
E6 13 SELMIN, O/AU  
E7 1 SELMINSKIKH N I/AU  
E8 4 SELMIRUBY S/AU  
E9 6 SELMISTRAITIS G/AU  
E10 1 SELMISTRAITIS G G/AU  
E11 37 SELMKE M/AU  
E12 4 SELMKE P/AU

=> s e3-4 or e6

L6 67 ("SELMIN O"/AU OR "SELMIN ORNELLA"/AU) OR  
"SELMIN, O"/AU

=> s l6 and protein?

2 FILES SEARCHED...

L7 43 L6 AND PROTEIN?

=> dup rem l7

DUPLICATE IS NOT AVAILABLE IN 'GENBANK'.  
ANSWERS FROM THESE FILES WILL BE CONSIDERED  
UNIQUE  
PROCESSING COMPLETED FOR L7  
L8 25 DUP REM L7 (18 DUPLICATES REMOVED)

=> d 1-10

L8 ANSWER 1 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 1  
AN 97:203523 BIOSIS  
DN 99502726  
TI Temporal regulation of the Dictyostelium glycogen  
phosphorylase 2  
gene.  
AU Rutherford C L; Selmin O; Peters-Weigel S  
CS Biol. Dep., Molecular and Cellular Biology Section, Virginia  
Polytechnic Inst. State Univ., Blacksburg, VA 24061, USA  
SO Biochimica et Biophysica Acta 1351 (1-2). 1997. 111-125.  
ISSN:  
0006-3002  
LA English

L8 ANSWER 2 OF 25 CAPLUS COPYRIGHT 1998 ACS  
AN 1997:85324 CAPLUS  
DN 126:114507  
TI Isolation and characterization of a novel gene induced by  
2,3,7,8-tetrachlorodibenzo-p-dioxin in rat liver  
AU Selmin, Ornella; Lucier, George W.; Clark, George C.;  
Tritscher, Angelika M.; Heuvel, John P. Vanden; Gastel, Jonathan  
A.;  
Walker, Nigel J.; Sutter, Thomas R.; Bell, Douglas A.  
CS Lab. Computational Biology, National Inst. Environ. Health  
Sciences,  
Research Triangle Park, NC, 27709, USA  
SO Carcinogenesis (1996), 17(12), 2609-2615  
CODEN: CRNGDP; ISSN: 0143-3334

PB Oxford University Press  
DT Journal  
LA English

L8 ANSWER 3 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 2  
AN 93:523019 BIOSIS  
DN BA96:136426  
TI REGULATION OF THE DICTYOSTELIUM GLYCOGEN  
PHOSPHORYLASE 2 GENE BY  
CYCLIC AMP.  
AU SUCIC J F; SELMIN O; RUTHERFORD C L  
CS BIOL. DEP., CELLULAR MOLECULAR BIOL. SECT.,  
VIRGINIA POLYTECHNIC  
INST., STATE UNIV., BLACKSBURG, VA 24061.  
SO DEV GENET 14 (4). 1993. 313-322. CODEN: DGNTDW  
ISSN: 0192-253X  
LA English

L8 ANSWER 4 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 3  
AN 92:147570 BIOSIS  
DN BA93:81795  
TI CLONING STRUCTURAL ANALYSIS AND EXPRESSION  
OF THE GLYCOGEN  
PHOSPHORYLASE-2 GENE IN DICTYOSTELIUM.  
AU RUTHERFORD C L; PEERY R B; SUCIC J F; YIN Y;  
ROGERS P V; LUO S;  
SELMIN O  
CS BIOL. DEP., VIRGINIA POLYTECHNIC INST. STATE  
UNIVERSITY, BLACKSBURG,  
VA. 24601.  
SO J BIOL CHEM 267 (4). 1992. 2294-2302. CODEN: JBCHA3  
ISSN: 0021-9258  
LA English

L8 ANSWER 5 OF 25 MEDLINE  
AN 90243721 MEDLINE  
DN 90243721  
TI The structure of anchorin CII, a collagen binding protein  
isolated from chondrocyte membrane.  
AU Fernandez M P; Selmin O; Martin G R; Yamada Y; Pfaffle M;  
Deutzmann R; Mollenhauer J; von der Mark K  
SO JOURNAL OF BIOLOGICAL CHEMISTRY, (1990 May 15)  
265 (14) 8344.  
Journal code: HIV. ISSN: 0021-9258.  
CY United States  
DT Journal; Article; (JOURNAL ARTICLE)  
LA English  
FS Priority Journals; Cancer Journals  
OS GENBANK-M30971  
EM 199008

L8 ANSWER 6 OF 25 CAPLUS COPYRIGHT 1998 ACS  
AN 1990:454611 CAPLUS  
DN 113:54611  
TI The structure anchorin CII, a collagen binding protein  
isolated from chondrocyte membrane [Erratum to document cited  
in  
CA109(1):2665p]  
AU Fernandez, M. Pilar; Selmin, Ornella; Martin, George R.;  
Yamada, Yoshihiko; Pfaffle, Michael; Deutzmann, Rainer;  
Mollenhauer, Jürgen; Von der Mark, Klaus  
CS Lab. Dev. Biol. Anomalies, Natl. Inst. Dent. Res., Bethesda,  
MD,  
20892, USA  
SO J. Biol. Chem. (1990), 265(14), 8344  
CODEN: JBCHA3; ISSN: 0021-9258  
DT Journal  
LA English

L8 ANSWER 7 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 4  
AN 90:485663 BIOSIS  
DN BR39:109684  
TI ANCHORIN CII A COLLAGEN-BINDING CHONDROCYTE  
SURFACE PROTEIN  
OF THE CALPACTIN FAMILY.

AU PFAEFFLE M; BORCHERT M; DEUTZMANN R; VON  
DER MARK K; FERNANDEZ M P;  
SELMIN O; YAMADA Y; MARTIN G; RUGGIERO F;  
GARRONE R  
CS MPG, ARBEITSGRUPPEN RHEUMATOL., D-8520  
ERLANGEN, FRG.  
SO MELLI, M. AND L. PARENTE (ED.). PROGRESS IN  
CLINICAL AND BIOLOGICAL  
RESEARCH, VOL. 349. CYTOKINES AND LIPOCORTINS  
IN INFLAMMATION AND  
DIFFERENTIATION; INTERNATIONAL CONFERENCE ON  
MOLECULAR AND CELLULAR  
BIOLOGY OF IL-1, TNF, AND LIPOCORTINS IN  
INFLAMMATION AND  
DIFFERENTIATION, SIENA, ITALY, OCTOBER 22-25, 1989.  
XX+463P.

WILEY-LISS: NEW YORK, NEW YORK, USA;  
CHICHESTER, ENGLAND, UK. ILLUS.  
0 (0). 1990. 147-158. CODEN: PCBRD2 ISBN: 0-471-56812-0  
ISSN:  
0361-7742  
DT Conference  
LA English

L8 ANSWER 8 OF 25 MEDLINE

AN 90377944 MEDLINE  
DN 90377944

TI Anchorin CII, a collagen-binding chondrocyte surface protein  
of the calpactin family.

AU Pfaffe M; Borchert M; Deutzmann R; von der Mark K;  
Fernandez M P;

Selmin O; Yamada Y; Martin G; Ruggiero F; Garrone R  
CS MPG, Arbeitsgruppen für Rheumatologie, Erlangen, FRG..  
SO PROGRESS IN CLINICAL AND BIOLOGICAL  
RESEARCH, (1990) 349 147-57.

Ref: 16

Journal code: PZ5. ISSN: 0361-7742.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

General Review; (REVIEW)  
(REVIEW, TUTORIAL)

LA English

FS Priority Journals

EM 199012

L8 ANSWER 9 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 5

AN 88:283835 BIOSIS

DN BA86:12102

TI THE STRUCTURE OF ANCHORIN CII A COLLAGEN  
BINDING PROTEIN

ISOLATED FROM CHONDROCYTE MEMBRANE.

AU FERNANDEZ M P; SELMIN O; MARTIN G R; YAMADA  
Y; PFAEFFLE M;

DEUTZMANN R; MOLLENHAUER J; VON DER MARK K  
CS LAB. DEVELOPMENTAL BIOLOGY AND ANOMALIES,  
NATL. INST. DENTAL RES.,

NATL. INST. HEALTH, BETHESDA, MD. 20892.

SO J BIOL CHEM 263 (12). 1988. 5921-5925. CODEN:  
JBCHA3 ISSN: 0021-9258

LA English

L8 ANSWER 10 OF 25 BIOSIS COPYRIGHT 1998 BIOSIS  
DUPLICATE 6

AN 89:90679 BIOSIS

DN BA87:44815

TI BIOSYNTHESIS SECRETION AND EXTRACELLULAR  
LOCALIZATION OF ANCHORIN CII

A COLLAGEN-BINDING PROTEIN OF THE CALPACTIN  
FAMILY.

AU PFAEFFLE M; RUGGIERO F; HOFMANN H; FERNANDEZ  
M P; SELMIN O;

YAMADA Y; GARRONE R; VON DER MARK K  
CS MAX-PLANCK-INST. BIOCHEM., DEP. CONNECTIVE  
TISSUE RES., D-8033

MARTINSRIED FRG.

SO EMBO (EUR MOL BIOL ORGAN) J 7 (8). 1988. 2335-2342.

CODEN: EMJODG

ISSN: 0261-4189

LA English

=> d 11-20

L8 ANSWER 11 OF 25 SCISEARCH COPYRIGHT 1998 ISI  
(R)

AN 88:420590 SCISEARCH

GA The Genuine Article (R) Number: P4640

TI BIOSYNTHESIS, SECRETION AND EXTRACELLULAR  
LOCALIZATION OF

ANCHORIN-C2, A COLLAGEN-BINDING PROTEIN OF  
THE CALPACTIN  
FAMILY

AU PFAEFFLE M (Reprint); RUGGIERO F; HOFMANN H;  
FERNANDEZ M P;

SELMIN O; YAMADA Y; GARRONE R; VONDERMARK K  
CS MAX PLANCK INST BIOCHEM, DEPT CONNECT  
TISSUE RES, D-8033

MARTINSRIED, FED REP GER (Reprint); UNIV LYON 1,  
HISTOL EXPTL LAB,

CNRS UA 244, F-69621 VILLEURBANNE, FRANCE; NIDR,  
DEV BIOL &

ANOMALIES LAB, BETHESDA, MD, 20205; MAX PLANCK  
GESELL, ARBEITSGRP

RHEUMATOL KLIN, D-8520 ERLANGEN, FED REP GER  
CYA GERMANY; FRANCE; USA

SO EMBO JOURNAL, (1988) Vol. 7, No. 8, pp. 2335-2342.

DT Article; Journal

FS LIFE

LA ENGLISH

REC Reference Count: 43

L8 ANSWER 12 OF 25 MEDLINE

DUPLICATE

AN 89209383 MEDLINE

DN 89209383

TI Glycogen phosphorylase in Dictyostelium discoideum:  
demonstration of

two developmentally regulated forms, purification to  
homogeneity,

immunochemical analysis, cAMP induction, in vitro translation,  
and

molecular cloning.

AU Rutherford C L; Naranan V; Brickey D A; Sucic J F; Rogers P  
V;

Selmin O

CS Biology Dept., Virginia Tech University, Blacksburg 24061.

NC AG00677 (NIA)

SO DEVELOPMENTAL GENETICS, (1988) 9 (4-5) 469-81.

Journal code: DEG. ISSN: 0192-253X.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English  
FS Priority Journals  
EM 198908

L8 ANSWER 13 OF 25 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): RNU63315 GenBank (R)  
GenBank ACC. NO. (GBN): U63315  
CAS REGISTRY NO. (RN): 180568-28-9  
SEQUENCE LENGTH (SQL): 1885  
MOLECULE TYPE (CI): mRNA; linear  
DIVISION CODE (CI): Rodents  
DATE (DATE): 2 Sep 1996  
DEFINITION (DEF): Rattus norvegicus 25-Dx (25Dx) mRNA,  
complete

cds.

SOURCE: Norway rat.  
ORGANISM (ORGN): Rattus norvegicus  
Eukaryotae; mitochondrial eukaryotes; Metazoa;  
Chordata; Vertebrata; Eutheria; Rodentia;  
Sciurognathi; Myomorpha; Muridae; Murinae; Rattus  
NUCLEIC ACID COUNT (NA): 550 a 399 c 427 g 509 t  
REFERENCE: 1 (bases 1 to 1885)

AUTHOR (AU): Selmin,O.; Lucier,G.; Clark,G.;  
Tritscher,A.; Vanden-Heuvel,J.; Gastel,J.;  
Walker,N.; Sutter,T.; Bell,D.A.

TITLE (TI): Isolation and characterization of a novel gene  
induced by 2,3,7,8 TCDD in rat liver

JOURNAL (SO): Carcinogenesis (1996) In press  
REFERENCE: 2 (bases 1 to 1885)

AUTHOR (AU): Bell,D.A.; Selmin,O.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (09-JUL-1996) LCBRA, NIEHS,  
P.O. Box

12233, Research Triangle Park, NC 27709, USA

L8 ANSWER 14 OF 25 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): GDANX5A10 GenBank (R)  
GenBank ACC. NO. (GBN): U01680  
CAS REGISTRY NO. (RN): 152001-35-9  
SEQUENCE LENGTH (SQL): 715  
MOLECULE TYPE (CI): DNA; linear  
DIVISION CODE (CI): Other vertebrates  
DATE (DATE): 7 Jan 1997  
DEFINITION (DEF): Gallus domesticus White Leghorn  
anchorin CII

(ANX5) gene, exon 13, 3'-untranslated region, and  
complete cds.

SEGMENT: 10 of 10

SOURCE: chicken.

ORGANISM (ORGN): Gallus gallus  
Eukaryotae; mitochondrial eukaryotes; Metazoa;  
Chordata; Vertebrata; Archosauria; Aves;  
Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 200 a 140 c 134 g 241 t  
REFERENCE: 1 (bases 1 to 715)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
Yamada,Y.; Pfaeffe,M.; Deutzmann,R.;  
Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

protein isolated from chondrocyte  
membrane

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
OTHER SOURCE (OS): CA 109:2665  
REFERENCE: 2 (bases 1 to 715)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
Yamada,Y.; Pfaeffe,M.; Deutzmann,R.;  
Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

protein isolated from chondrocyte  
membrane

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
OTHER SOURCE (OS): CA 113:54611

REFERENCE: 3 (bases 1 to 715)

AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII  
(chick annexin V)

JOURNAL (SO): Gene, 141, 179-186 (1994)

OTHER SOURCE (OS): CA 121:51129

REFERENCE: 4 (bases 1 to 715)

AUTHOR (AU): Fernandez,M.P.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
Fernandez,

Universidad de Oviedo, Departamento de Biologia  
Funcional, c/Julian Claveria, 33071 Oviedo,  
Asturias, Spain

L8 ANSWER 15 OF 25 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): GDANX5A09 GenBank (R)  
GenBank ACC. NO. (GBN): U01679  
CAS REGISTRY NO. (RN): 152001-34-8  
SEQUENCE LENGTH (SQL): 773  
MOLECULE TYPE (CI): DNA; linear  
DIVISION CODE (CI): Other vertebrates  
DATE (DATE): 7 Jan 1997  
DEFINITION (DEF): Gallus domesticus White Leghorn  
anchorin CII exon  
12.

SEGMENT: 9 of 10

SOURCE: chicken.

ORGANISM (ORGN): Gallus gallus  
Eukaryotae; mitochondrial eukaryotes; Metazoa;  
Chordata; Vertebrata; Archosauria; Aves;  
Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 209 a 125 c 189 g 250 t  
REFERENCE: 1 (bases 1 to 773)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
Yamada,Y.; Pfaeffe,M.; Deutzmann,R.;  
Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

protein isolated from chondrocyte  
membrane

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
OTHER SOURCE (OS): CA 109:2665

REFERENCE: 2 (bases 1 to 773)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
Yamada,Y.; Pfaeffe,M.; Deutzmann,R.;  
Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

protein isolated from chondrocyte  
membrane

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
OTHER SOURCE (OS): CA 113:54611

REFERENCE: 3 (bases 1 to 773)

AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII

(chick annexin V)  
 JOURNAL (SO): Gene, 141, 179-186 (1994)  
 OTHER SOURCE (OS): CA 121:51129  
 REFERENCE: 4 (bases 1 to 773)  
 AUTHOR (AU): Fernandez, M.P.  
 TITLE (TI): Direct Submission  
 JOURNAL (SO): Submitted (14-SEP-1993) Maria P. Fernandez,  
 Universidad de Oviedo, Departamento de Biología  
 Funcional, c/Julian Claveria, 33071 Oviedo,  
 Asturias, Spain

L8 ANSWER 16 OF 25 GENBANK.RTM. COPYRIGHT  
 1998

LOCUS (LOC): GDANX5A08 GenBank (R)  
 GenBank ACC. NO. (GBN): U01678  
 CAS REGISTRY NO. (RN): 152001-33-7  
 SEQUENCE LENGTH (SQL): 735  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn  
 anchorin CII  
 (ANX5) gene, exons 10 and 11.  
 SEGMENT: 8 of 10  
 SOURCE: chicken.  
 ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus  
 NUCLEIC ACID COUNT (NA): 195 a 129 c 155 g 256 t  
 REFERENCE: 1 (bases 1 to 735)  
 AUTHOR (AU): Fernandez, M.P.; Selmin, O.; Martin, G.R.;  
 Yamada, Y.; Pfaffle, M.; Deutzmann, R.;  
 Mollenhauer, J.; von der Mark, K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

**protein isolated from chondrocyte  
 membrane**  
 JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665  
 REFERENCE: 2 (bases 1 to 735)  
 AUTHOR (AU): Fernandez, M.P.; Selmin, O.; Martin, G.R.;  
 Yamada, Y.; Pfaffle, M.; Deutzmann, R.;  
 Mollenhauer, J.; von der Mark, K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

**protein isolated from chondrocyte  
 membrane**  
 JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
 OTHER SOURCE (OS): CA 113:54611  
 REFERENCE: 3 (bases 1 to 735)  
 AUTHOR (AU): Fernandez, M.P.; Fernandez, M.R.;  
 Morgan, R.O.  
 TITLE (TI): Structure of the gene encoding anchorin CII  
 (chick annexin V)  
 JOURNAL (SO): Gene, 141, 179-186 (1994)  
 OTHER SOURCE (OS): CA 121:51129  
 REFERENCE: 4 (bases 1 to 735)  
 AUTHOR (AU): Fernandez, M.P.  
 TITLE (TI): Direct Submission  
 JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
 Fernandez,

Universidad de Oviedo, Departamento de Biología  
 Funcional, c/Julian Claveria, 33071 Oviedo,  
 Asturias, Spain

L8 ANSWER 17 OF 25 GENBANK.RTM. COPYRIGHT  
 1998

LOCUS (LOC): GDANX5A07 GenBank (R)  
 GenBank ACC. NO. (GBN): U01677  
 CAS REGISTRY NO. (RN): 152001-32-6  
 SEQUENCE LENGTH (SQL): 313  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn  
 anchorin CII  
 (ANX5) gene, exon 9.  
 SEGMENT: 7 of 10  
 SOURCE: chicken.  
 ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus  
 NUCLEIC ACID COUNT (NA): 88 a 55 c 66 g 104 t  
 REFERENCE: 1 (bases 1 to 313)  
 AUTHOR (AU): Fernandez, M.P.; Selmin, O.; Martin, G.R.;  
 Yamada, Y.; Pfaffle, M.; Deutzmann, R.;  
 Mollenhauer, J.; von der Mark, K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

**protein isolated from chondrocyte  
 membrane**  
 JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665  
 REFERENCE: 2 (bases 1 to 313)  
 AUTHOR (AU): Fernandez, M.P.; Selmin, O.; Martin, G.R.;  
 Yamada, Y.; Pfaffle, M.; Deutzmann, R.;  
 Mollenhauer, J.; von der Mark, K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

**protein isolated from chondrocyte  
 membrane**  
 JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
 OTHER SOURCE (OS): CA 113:54611  
 REFERENCE: 3 (bases 1 to 313)  
 AUTHOR (AU): Fernandez, M.P.; Fernandez, M.R.;  
 Morgan, R.O.  
 TITLE (TI): Structure of the gene encoding anchorin CII  
 (chick annexin V)  
 JOURNAL (SO): Gene, 141, 179-186 (1994)  
 OTHER SOURCE (OS): CA 121:51129  
 REFERENCE: 4 (bases 1 to 313)  
 AUTHOR (AU): Fernandez, M.P.  
 TITLE (TI): Direct Submission  
 JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
 Fernandez,

Universidad de Oviedo, Departamento de Biología  
 Funcional, c/Julian Claveria, 33071 Oviedo,  
 Asturias, Spain

L8 ANSWER 18 OF 25 GENBANK.RTM. COPYRIGHT  
 1998

LOCUS (LOC): GDANX5A06 GenBank (R)  
 GenBank ACC. NO. (GBN): U01676  
 CAS REGISTRY NO. (RN): 151999-85-8  
 SEQUENCE LENGTH (SQL): 639  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn  
 anchorin CII

(ANX5) gene, exons 7 and 8.  
 SEGMENT: 6 of 10  
 SOURCE: chicken.  
 ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus  
 NUCLEIC ACID COUNT (NA): 164 a 105 c 161 g 209 t  
 REFERENCE: 1 (bases 1 to 639)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte  
 membrane  
 JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665  
 REFERENCE: 2 (bases 1 to 639)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte  
 membrane  
 JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
 OTHER SOURCE (OS): CA 113:54611  
 REFERENCE: 3 (bases 1 to 639)  
 AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
 Morgan,R.O.  
 TITLE (TI): Structure of the gene encoding anchorin CII  
 (chick annexin V)  
 JOURNAL (SO): Gene, 141, 179-186 (1994)  
 OTHER SOURCE (OS): CA 121:51129  
 REFERENCE: 4 (bases 1 to 639)  
 AUTHOR (AU): Fernandez,M.P.  
 TITLE (TI): Direct Submission  
 JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
 Fernandez,  
 Universidad de Oviedo, Departamento de Biologia  
 Funcional, c/Julian Claveria, 33071 Oviedo,  
 Asturias, Spain

L8 ANSWER 19 OF 25 GENBANK.RTM. COPYRIGHT  
 1998

LOCUS (LOC): GDANX5A05 GenBank (R)  
 GenBank ACC. NO. (GBN): U01675  
 CAS REGISTRY NO. (RN): 151999-84-7  
 SEQUENCE LENGTH (SQL): 418  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn  
 anchorin CII

(ANX5) gene, exon 6.  
 SEGMENT: 5 of 10  
 SOURCE: chicken.  
 ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus  
 NUCLEIC ACID COUNT (NA): 134 a 69 c 84 g 131 t  
 REFERENCE: 1 (bases 1 to 418)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte  
 membrane  
 JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665  
 REFERENCE: 2 (bases 1 to 418)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte  
 membrane  
 JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
 OTHER SOURCE (OS): CA 113:54611  
 REFERENCE: 3 (bases 1 to 418)  
 AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
 Morgan,R.O.  
 TITLE (TI): Structure of the gene encoding anchorin CII  
 (chick annexin V)  
 JOURNAL (SO): Gene, 141, 179-186 (1994)  
 OTHER SOURCE (OS): CA 121:51129  
 REFERENCE: 4 (bases 1 to 418)  
 AUTHOR (AU): Fernandez,M.P.  
 TITLE (TI): Direct Submission  
 JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
 Fernandez,  
 Universidad de Oviedo, Departamento de Biologia  
 Funcional, c/Julian Claveria, 33071 Oviedo,  
 Asturias, Spain

L8 ANSWER 20 OF 25 GENBANK.RTM. COPYRIGHT  
 1998

LOCUS (LOC): GDANX5A04 GenBank (R)  
 GenBank ACC. NO. (GBN): U01674  
 CAS REGISTRY NO. (RN): 151999-96-1  
 SEQUENCE LENGTH (SQL): 406  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn  
 anchorin CII

(ANX5) gene, exon 5.  
 SEGMENT: 4 of 10  
 SOURCE: chicken.  
 ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus  
 NUCLEIC ACID COUNT (NA): 130 a 57 c 81 g 138 t  
 REFERENCE: 1 (bases 1 to 406)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte  
 membrane  
 JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665  
 REFERENCE: 2 (bases 1 to 406)  
 AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.  
 TITLE (TI): The structure of anchorin CII, a collagen  
 binding

protein isolated from chondrocyte

membrane  
JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
OTHER SOURCE (OS): CA 113:54611  
REFERENCE: 3 (bases 1 to 406)  
AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII  
(chick annexin V)

JOURNAL (SO): Gene, 141, 179-186 (1994)

OTHER SOURCE (OS): CA 121:51129

REFERENCE: 4 (bases 1 to 406)

AUTHOR (AU): Fernandez,M.P.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
Fernandez,

Universidad de Oviedo, Departamento de Biología  
Funcional, c/Julian Claveria, 33071 Oviedo,  
Asturias, Spain

=> d 21-25

L8 ANSWER 21 OF 25 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): GDANX5A03 GenBank (R)

GenBank ACC. NO. (GBN): U01673

CAS REGISTRY NO. (RN): 151999-95-0

SEQUENCE LENGTH (SQL): 176

MOLECULE TYPE (CI): DNA; linear

DIVISION CODE (CI): Other vertebrates

DATE (DATE): 7 Jan 1997

DEFINITION (DEF): Gallus domesticus White Leghorn  
anchorin CII

(ANX5) gene, exon 4.

SEGMENT: 3 of 10

SOURCE: chicken.

ORGANISM (ORGN): Gallus gallus

Eukaryotae; mitochondrial eukaryotes; Metazoa;

Chordata; Vertebrata; Archosauria; Aves;

Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 55 a 29 c 37 g 55 t

REFERENCE: 1 (bases 1 to 176)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;

Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;

Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

**protein isolated from chondrocyte  
membrane**

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)

OTHER SOURCE (OS): CA 109:2665

REFERENCE: 2 (bases 1 to 176)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;

Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;

Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

**protein isolated from chondrocyte  
membrane**

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)

OTHER SOURCE (OS): CA 113:54611

REFERENCE: 3 (bases 1 to 176)

AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;

Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII  
(chick annexin V)

JOURNAL (SO): Gene, 141, 179-186 (1994)

OTHER SOURCE (OS): CA 121:51129

REFERENCE: 4 (bases 1 to 176)

AUTHOR (AU): Fernandez,M.P.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
Fernandez,

Universidad de Oviedo, Departamento de Biología  
Funcional, c/Julian Claveria, 33071 Oviedo,  
Asturias, Spain

L8 ANSWER 22 OF 25 GENBANK.RTM. COPYRIGHT  
1998

LOCUS (LOC): GDANX5A02 GenBank (R)

GenBank ACC. NO. (GBN): U01672

CAS REGISTRY NO. (RN): 151999-94-9

SEQUENCE LENGTH (SQL): 285

MOLECULE TYPE (CI): DNA; linear

DIVISION CODE (CI): Other vertebrates

DATE (DATE): 7 Jan 1997

DEFINITION (DEF): Gallus domesticus White Leghorn  
anchorin CII

(ANX5) gene, exon 3.

SEGMENT: 2 of 10

SOURCE: chicken.

ORGANISM (ORGN): Gallus gallus

Eukaryotae; mitochondrial eukaryotes; Metazoa;

Chordata; Vertebrata; Archosauria; Aves;

Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 77 a 59 c 59 g 90 t

REFERENCE: 1 (bases 1 to 285)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;

Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;

Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

**protein isolated from chondrocyte  
membrane**

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)

OTHER SOURCE (OS): CA 109:2665

REFERENCE: 2 (bases 1 to 285)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;

Yamada,Y.; Pfaeffle,M.; Deutzmann,R.;

Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen  
binding

**protein isolated from chondrocyte  
membrane**

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)

OTHER SOURCE (OS): CA 113:54611

REFERENCE: 3 (bases 1 to 285)

AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;

Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII  
(chick annexin V)

JOURNAL (SO): Gene, 141, 179-186 (1994)

OTHER SOURCE (OS): CA 121:51129

REFERENCE: 4 (bases 1 to 285)

AUTHOR (AU): Fernandez,M.P.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (14-SEP-1993) Maria P.  
Fernandez,

Universidad de Oviedo, Departamento de Biología  
Funcional, c/Julian Claveria, 33071 Oviedo,  
Asturias, Spain

L8 ANSWER 23 OF 25 GENBANK.RTM. COPYRIGHT  
1998



LOCUS (LOC): GDANX5A01 GenBank (R)  
 GenBank ACC. NO. (GBN): U01671  
 CAS REGISTRY NO. (RN): 151999-93-8  
 SEQUENCE LENGTH (SQL): 1777  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Other vertebrates  
 DATE (DATE): 7 Jan 1997  
 DEFINITION (DEF): Gallus domesticus White Leghorn anchorin CII

(ANX5) gene, 5'-upstream region, exons 1 and 2.  
 SEGMENT: 1 of 10  
 SOURCE: chicken.

ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 388 a 458 c 501 g 430 t  
 REFERENCE: 1 (bases 1 to 1777)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen binding

protein isolated from chondrocyte membrane

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)  
 OTHER SOURCE (OS): CA 109:2665

REFERENCE: 2 (bases 1 to 1777)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen binding

protein isolated from chondrocyte membrane

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)  
 OTHER SOURCE (OS): CA 113:54611

REFERENCE: 3 (bases 1 to 1777)

AUTHOR (AU): Fernandez,M.P.; Fernandez,M.R.;  
 Morgan,R.O.

TITLE (TI): Structure of the gene encoding anchorin CII (chick annexin V)

JOURNAL (SO): Gene, 141, 179-186 (1994)

OTHER SOURCE (OS): CA 121:51129

REFERENCE: 4 (bases 1 to 1777)

AUTHOR (AU): Fernandez,M.P.

TITLE (TI): Direct Submission

JOURNAL (SO): Submitted (14-SEP-1993) Maria P. Fernandez,

Universidad de Oviedo, Departamento de Biologia Funcional, c/Julian Claveria, 33071 Oviedo, Asturias, Spain

L8 ANSWER 24 OF 25 GENBANK.RTM. COPYRIGHT 1998

LOCUS (LOC): DDIGP2A GenBank (R)  
 GenBank ACC. NO. (GBN): M77492  
 CAS REGISTRY NO. (RN): 140345-55-7  
 SEQUENCE LENGTH (SQL): 3594  
 MOLECULE TYPE (CI): DNA; linear  
 DIVISION CODE (CI): Plants, fungi, algae  
 DATE (DATE): 14 Jul 1992  
 DEFINITION (DEF): Dictyostelium discoideum glycoprotein phosphorylase 2 (glpD) gene, complete cds.  
 KEYWORDS (ST): cytoplasmic protein; glycogen phosphorylase 2  
 SOURCE: Dictyostelium discoideum (strain AX-3) DNA.

ORGANISM (ORGN): Dictyostelium discoideum  
 Eukaryotae; mitochondrial eukaryotes;  
 Dictyosteliida; Dictyostelium  
 NUCLEIC ACID COUNT (NA): 1304 a 501 c 550 g 1239 t  
 REFERENCE: 1 (bases 1 to 3594)  
 AUTHOR (AU): Rutherford,C.L.; Peery,R.B.; Sucic,J.F.; Yin,Y.;

Rogers,P.V.; Luo,S.; Selmin,O.

TITLE (TI): cloning, structural analysis, and expression of the glycogen phosphorylase-2 gene in Dictyostelium

JOURNAL (SO): J. Biol. Chem., 267, 2294-2302 (1992)

OTHER SOURCE (OS): CA 118:164065

L8 ANSWER 25 OF 25 GENBANK.RTM. COPYRIGHT 1998

LOCUS (LOC): CHKANCC2A GenBank (R)  
 GenBank ACC. NO. (GBN): M30971 J03194

CAS REGISTRY NO. (RN): 140313-10-6

SEQUENCE LENGTH (SQL): 1229

MOLECULE TYPE (CI): mRNA; linear

DIVISION CODE (CI): Other vertebrates

DATE (DATE): 15 Sep 1990

DEFINITION (DEF): Chicken anchorin CII mRNA, 3' end.

KEYWORDS (ST): anchorin; collagen-binding protein

SOURCE: Chicken cartilage (sternum) and bone, cDNA to mRNA, clones A[1,4,6,7,14,15,22,23].

ORGANISM (ORGN): Gallus gallus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa;  
 Chordata; Vertebrata; Archosauria; Aves;  
 Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus

NUCLEIC ACID COUNT (NA): 353 a 230 c 310 g 336 t

ORIGIN: 28 bp upstream of AccI site.

REFERENCE: 1 (bases 1 to 1229)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen binding

protein isolated from chondrocyte membrane

JOURNAL (SO): J. Biol. Chem., 263, 5921-5925 (1988)

OTHER SOURCE (OS): CA 109:2665

REFERENCE: 2 (bases 373 to 504)

AUTHOR (AU): Fernandez,M.P.; Selmin,O.; Martin,G.R.;  
 Yamada,Y.; Pfaffle,M.; Deutzmann,R.;  
 Mollenhauer,J.; von der Mark,K.

TITLE (TI): The structure of anchorin CII, a collagen binding

protein isolated from chondrocyte membrane

JOURNAL (SO): J. Biol. Chem., 265, 8344-8344 (1990)

OTHER SOURCE (OS): CA 113:54611

=> e falkenstein e/au

E1 1 FALKENSTEIN D F/AU  
 E2 2 FALKENSTEIN DORIS/AU  
 E3 20 --> FALKENSTEIN E/AU  
 E4 2 FALKENSTEIN ELIEZER/AU  
 E5 3 FALKENSTEIN ELISABETH/AU  
 E6 1 FALKENSTEIN ELLIEZER/AU  
 E7 10 FALKENSTEIN G/AU  
 E8 1 FALKENSTEIN G C/AU  
 E9 1 FALKENSTEIN G G/AU  
 E10 1 FALKENSTEIN G L/AU  
 E11 2 FALKENSTEIN GARY L/AU

E12 22 FALKENSTEIN GEORG/AU

=> s e5

L9 3 "FALKENSTEIN ELISABETH"/AU

=> dup rem l9

DUPLICATE IS NOT AVAILABLE IN 'GENBANK'.  
ANSWERS FROM THESE FILES WILL BE CONSIDERED  
UNIQUE  
PROCESSING COMPLETED FOR L9  
L10 3 DUP REM L9 (0 DUPLICATES REMOVED)

=> d l-3

L10 ANSWER 1 OF 3 CAPLUS COPYRIGHT 1998 ACS  
AN 1996:729806 CAPLUS  
DN 126:42820  
TI Full-length cDNA sequence of a progesterone membrane-binding  
protein  
from porcine vascular smooth muscle cells  
AU **Falkenstein, Elisabeth**; Meyer, Christiane; Eisen,  
Christoph; Scriba, Peter C.; Wehling, Martin  
CS Div. Clinical Pharmacology, Medizinische Klinik, Klinikum  
Innenstadt, Univ. Munich, Munich, 80336, Germany  
SO Biochem. Biophys. Res. Commun. (1996), 229(1), 86-89  
CODEN: BBRC A9; ISSN: 0006-291X  
PB Academic  
DT Journal  
LA English

L10 ANSWER 2 OF 3 CAPLUS COPYRIGHT 1998 ACS  
AN 1994:526521 CAPLUS  
DN 121:126521  
TI Full-length cDNA sequences for both ferredoxin-thioredoxin  
reductase  
subunits from spinach (*Spinacia oleracea* L.)  
AU **Falkenstein, Elisabeth**; von Schaewen, Antje; Scheibe,  
Renate  
CS Pflanzenphysiologie, FB 5 Biologie/Chemie, Universitaet  
Osnabrueck,  
Osnabrueck, D-49069, Germany  
SO Biochim. Biophys. Acta (1994), 1185(2), 252-4  
CODEN: BBACAQ; ISSN: 0006-3002  
DT Journal  
LA English

L10 ANSWER 3 OF 3 CAPLUS COPYRIGHT 1998 ACS  
AN 1991:652304 CAPLUS  
DN 115:252304  
TI Methyljasmonate and .alpha.-linolenic acid are potent inducers  
of  
tendrill coiling  
AU **Falkenstein, Elisabeth**; Groth, Beate; Mithoefer, Axel;  
Weiler, Elmar W.  
CS Ruhr-Univ., Bochum, W-4630, Fed. Rep. Ger.  
SO Planta (1991), 185(3), 316-22  
CODEN: PLANAB; ISSN: 0032-0935  
DT Journal  
LA English

=> e goli s/au

E1 2 GOLI P/AU  
E2 1 GOLI R/AU  
E3 0 --> GOLI S/AU  
E4 1 GOLI S M/AU  
E5 1 GOLI S R/AU

E6 1 GOLI SASANK MOHAN/AU  
E7 1 GOLI SURYA K/AU  
E8 1 GOLI T/AU  
E9 24 GOLI U B/AU  
E10 8 GOLI UMESH B/AU  
E11 1 GOLI UMESH BALKRISHNA/AU  
E12 48 GOLI V/AU

=> s e7

L11 1 "GOLI SURYA K"/AU

=> d

L11 ANSWER 1 OF 1 CAPLUS COPYRIGHT 1998 ACS  
AN 1997:719632 CAPLUS  
DN 128:1462  
TI Human phosphorylase kinase .gamma. subunit  
IN Bandman, Olga; Goli, Surya K.  
PA Incyte Pharmaceuticals, Inc., USA  
SO U.S., 25 pp.  
CODEN: USXXAM  
PI US 5683910 A 971104  
AI US 96-713828 960913  
DT Patent  
LA English

\*\*\*\*\*  
 \* WELCOME TO THE \*  
 \* U.S. PATENT TEXT FILE \*  
 \*\*\*\*\*

=> s systar?

L1 1 SYSTAR?

=> d

1. 5,581,482, Dec. 3, 1996, Performance monitor for digital computer system; Gregory B. Wiedenman, et al., 364/551.01, 550; 395/184.01 :IMAGE AVAILABLE:

=> del 11

DELETE L1? (Y/N):y

=> s cystar? and cytokine?

3 CYSTAR?

2790 CYTOKINE?

L1 0 CYSTAR? AND CYTOKINE?

=> s (cytokine and steroid?)(5a)receptor?

PROXIMITY OPERATION NOT ALLOWED

=> s (cytokine and steroid?)

1493 CYTOKINE

13685 STEROID?

L2 309 (CYTOKINE AND STEROID?)

=> d 1-10

1. 5,708,158, Jan. 13, 1998, Nuclear factors and binding assays; Timothy Hoey, 536/23.5, 23.1 :IMAGE AVAILABLE:

2. 5,708,147, Jan. 13, 1998, Mononuclear leukocyte directed endothelial adhesion molecule associated with atherosclerosis; Michael A. Gimbrone, Jr., et al., 530/388.7; 436/63, 86; 530/350, 395 :IMAGE AVAILABLE:

3. 5,708,025, Jan. 13, 1998, Methods for promoting wound healing; Dvorit Samid, 514/538, 563, 567, 885, 886, 928 :IMAGE AVAILABLE:

4. 5,707,821, Jan. 13, 1998, Identification of phospholipase A2 inhibitors in A.beta. peptide-mediated neurodegenerative disease; Russell E. Rydel, et al., 435/18, 4; 514/12 :IMAGE AVAILABLE:

5. 5,707,621, Jan. 13, 1998, Suppression of nephritis-induced protein excretion by anti-IL-8; Kouji Matsushima, 424/145.1, 130.1, 133.1, 158.1, 809; 435/325, 326, 328, 335; 530/387.1, 387.3, 388.1, 388.23 :IMAGE AVAILABLE:

6. 5,705,351, Jan. 6, 1998, Diagnosis of cancer using tumor-mimetic cell surface antigen from chemically modified normal cells; Madhao B. Sahasrabudhe, 435/7.23, 7.1, 7.2, 7.24, 961; 436/63, 64, 813 :IMAGE AVAILABLE:

7. 5,705,349, Jan. 6, 1998, Methods for preparing polynucleotides encoding orphan receptor ligands; Richard D. Holly, et al., 435/7.2, 6,

7.21, 69.1, 69.5, 172.1, 372, 372.1, 405; 436/501; 536/23.1, 23.5 :IMAGE AVAILABLE:

8. 5,703,098, Dec. 30, 1997, Immunotherapeutic imides/amides; George W. Muller, et al., 514/339, 417; 546/277.1; 548/476 :IMAGE AVAILABLE:

9. 5,703,092, Dec. 30, 1997, Hydroxamic acid compounds as metalloprotease and TNF inhibitors; Chu-Biao Xue, et al., 514/303, 394, 399; 546/118, 334; 548/180, 204, 217, 235, 247, 253, 309.7, 338.1, 375.1; 562/623 :IMAGE AVAILABLE:

10. 5,703,060, Dec. 30, 1997, Uses of aloe products in the prevention and treatment of infections and infestations; Bill H. McAnalley, et al., 514/54, 885 :IMAGE AVAILABLE:

=> d 11-20

11. 5,703,055, Dec. 30, 1997, Generation of antibodies through lipid mediated DNA delivery; Philip L. Felgner, et al., 514/44; 424/130.1, 184.1; 435/69.3; 935/60, 65 :IMAGE AVAILABLE:

12. 5,703,048, Dec. 30, 1997, Protection against liver damage by HGF; Filip Roos, et al., 514/12; 435/360; 514/2, 838, 893, 894; 530/350, 399; 935/13 :IMAGE AVAILABLE:

13. 5,702,697, Dec. 30, 1997, Treatment for biological damage using a colony stimulating factor and a biological modifier; Robert Zimmerman, et al., 424/85.1; 514/2, 8, 885; 530/351 :IMAGE AVAILABLE:

14. 5,698,706, Dec. 16, 1997, Heterocyclic amides and methods of use; Andrew Douglas Baxter, et al., 548/314.7 :IMAGE AVAILABLE:

15. 5,698,579, Dec. 16, 1997, Cyclic amides; George W. Muller, 514/416; 548/512 :IMAGE AVAILABLE:

16. 5,698,399, Dec. 16, 1997, Detecting genetic predisposition for osteoporosis; Gordon W. Duff, et al., 435/6, 91.2 :IMAGE AVAILABLE:

17. 5,698,232, Dec. 16, 1997, Pharmaceutical composition for treatment of sudden deafness; Shinichi Kanemaru, et al., 514/2, 21, 46, 52, 59, 557 :IMAGE AVAILABLE:

18. 5,698,195, Dec. 16, 1997, Methods of treating rheumatoid arthritis using chimeric anti-TNF antibodies; Junming Le, et al., 424/133.1, 141.1, 142.1, 145.1; 514/825; 530/351, 387.3, 388.1, 388.23 :IMAGE AVAILABLE:

19. 5,698,178, Dec. 16, 1997, Polyspecific immunoconjugates and antibody composites for targeting the multidrug resistant phenotype; David M. Goldenberg, 424/1.49, 1.53, 9.341, 9.6 :IMAGE AVAILABLE:

20. 5,695,761, Dec. 9, 1997, Suppression of nitric oxide production by osteopontin; David T. Denhardt, et al., 424/184.1, 85.5, 278.1; 514/2, 12; 530/300, 326, 330, 351 :IMAGE AVAILABLE:

=> s (steroid? or progesterone?) and (membrane binding protein?)

13685 STEROID?  
2684 PROGESTERONE?  
80732 MEMBRANE  
95444 BINDING  
72438 PROTEIN?  
9 MEMBRANE BINDING PROTEIN?  
(MEMBRANE(W)BINDING(W)PROTEIN?)

L3 1 (STEROID? OR PROGESTERONE?) AND (MEMBRANE BINDING PROTEIN?)

=> d

1. 5,530,114, Jun. 25, 1996, Oligonucleotide modulation of arachidonic acid metabolism; Clarence F. Bennett, et al., 536/24.3; 435/6; 536/24.1 :IMAGE AVAILABLE:

=> d ab

US PAT NO: 5,530,114 :IMAGE AVAILABLE: L3: 1 of 1

ABSTRACT:

Compositions and methods are provided for the treatment and diagnosis of diseases amenable to modulation of the synthesis or metabolism of arachidonic acid and related compounds. In accordance with preferred embodiments, oligonucleotides and oligonucleotide analogs are provided which are specifically hybridizable with nucleic acids encoding 5-lipoxygenase, 5-lipoxygenase activating proteins, LTA.sub.4 hydrolase, phospholipase A.sub.2, phospholipase C, and coenzyme A-independent transacylase. The oligonucleotide comprises nucleotide units sufficient in identity and number to effect said specific hybridization. In other preferred embodiments, the oligonucleotides are specifically hybridizable with a transcription initiation site, a translation initiation site, and intron/exon junction. Methods of treating animals suffering from disease amenable to therapeutic intervention by modulating arachidonic acid synthesis or metabolism with an oligonucleotide or oligonucleotide analog specifically hybridizable with RNA or DNA corresponding to one of the foregoing proteins are disclosed. Methods for treatment of diseases responding to modulation of arachidonic acid synthesis or metabolism are disclosed.

=> d kwic

US PAT NO: 5,530,114 :IMAGE AVAILABLE: L3: 1 of 1

SUMMARY:

BSUM(37)

**\*\*Steroids\*\*** exhibiting glucocorticoid activity also exhibit anti-inflammatory activity, possibly by inhibiting the release of arachidonic acid from cell membranes. **\*\*Steroids\*\*** constitute one of the most widely prescribed classes of agents currently available. They are used to treat a variety of . . .

DETDESC:

DETD(13)

The . . . site. 5-LO contains 2 domains which show 50- to 60% homology to the 17 amino acid consensus sequence for calcium-dependent

**\*\*membrane\*\* \*\*binding\*\* \*\*proteins\*\*** such as lipocortin. The similarities between 5-LO and the calcium-dependent **\*\*membrane\*\* \*\*binding\*\* \*\*proteins\*\*** may explain the calcium-dependent translocation of 5-LO from the cytosol to membranes.

=> e selmin, o/in

E#	FILE	FREQUENCY	TERM
E1	USPAT	1	SELMEZCI, ANDRAS/IN
E2	USPAT	1	SELMI, FATHI/IN
E3	USPAT	0 -->	SELMIN, O/IN
E4	USPAT	1	SELMIN, PAOLO/IN
E5	USPAT	3	SELNA, ERICH/IN
E6	USPAT	4	SELNER, ALLEN J/IN
E7	USPAT	2	SELNER, MARC D/IN
E8	USPAT	1	SELNES, ARNOLD/IN
E9	USPAT	1	SELNESS, JERRY N/IN
E10	USPAT	26	SELNICK, HAROLD G/IN
E11	USPAT	4	SELNICK, LESTER L/IN
E12	USPAT	1	SELNICK, RANDY/IN

=> e falkenstein, e/in

E#	FILE	FREQUENCY	TERM
E1	USPAT	1	FALKENSTEIN, ALBERT J/IN
E2	USPAT	4	FALKENSTEIN, DON I/IN
E3	USPAT	0 -->	FALKENSTEIN, E/IN
E4	USPAT	1	FALKENSTEIN, ELIEZER/IN
E5	USPAT	11	FALKENSTEIN, GEORG/IN
E6	USPAT	1	FALKENSTEIN, HEINZ P/IN
E7	USPAT	1	FALKENSTEIN, HERMANN J/IN
E8	USPAT	1	FALKENSTEIN, KENT C/IN
E9	USPAT	1	FALKENSTEIN, ROGER/IN
E10	USPAT	2	FALKENSTEIN, RUDOLF/IN
E11	USPAT	1	FALKENSTROM, CHE HSIN/IN
E12	USPAT	1	FALKENSTROM, LEE J/IN

=> s e4

L4 1 "FALKENSTEIN, ELIEZER"/IN

=> d

1. 5,045,282, Sep. 3, 1991, Optical fiber sensing device for analysis; Amnon Kritzman, et al., 422/56; 250/227.14, 483.1; 422/57, 58, 82.06, 82.07, 82.08, 82.11; 436/164, 169, 172 :IMAGE AVAILABLE:

=> e goli, s/in

E#	FILE	FREQUENCY	TERM
E1	USPAT	1	GOLGER, LEONID I/IN
E2	USPAT	1	GOLI, MARIA E/IN
E3	USPAT	0 -->	GOLI, S/IN
E4	USPAT	1	GOLI, SURYA K/IN
E5	USPAT	2	GOLIA, DOMINICK JR/IN
E6	USPAT	3	GOLIA, KENNETH R/IN

E7 USPAT 1 GOLIA, MICHEL/IN  
E8 USPAT 2 GOLIA, SALVATORE/IN  
E9 USPAT 1 GOLIAN, TIMOTHY G/IN  
E10 USPAT 1 GOLIARD, FRANCOIS/IN  
E11 USPAT 1 GOLIAS, BERNARD J/IN  
E12 USPAT 8 GOLIAS, JOSEPH H/IN

=> s e4

L5 1 "GOLI, SURYA K"/IN

=> d

1. 5,683,910, Nov. 4, 1997, Human phosphorylase kinase gamma subunit;

Olga Bandman, et al., 435/194; 424/94.5; 435/69.1, 193, 252.3, 320.1;

530/350; 536/23.2 :IMAGE AVAILABLE:

=> d his

(FILE 'USPAT' ENTERED AT 17:44:05 ON 16 JAN 1998)

L1 0 S CYSTAR? AND CYTOKINE?  
L2 309 S (CYTOKINE AND STEROID?)  
L3 1 S (STEROID? OR PROGESTERONE?) AND  
(MEMBRANE BINDING PROTEI  
N?)

E SELMIN, O/IN

E FALKENSTEIN, E/IN

L4 1 S E4

E GOLI, S/IN

L5 1 S E4

\*\*\*\*\*STN Columbus\*\*\*\*\*

FILE 'HOME' ENTERED AT 22:16:55 ON 18 JAN 1998

=> index bioscience

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	
	ENTRY SESSION
FULL ESTIMATED COST	0.15 0.15

INDEX 'AGRICOLA, AIDSLINE, ANABSTR, AQUASCI,  
BIOBUSINESS, BIOSIS, BIOTECHABS,  
BIOTECHDS, CABA, CANCERLIT, CAPLUS, CEABA,  
CEN, CIN, CJACS, CJELSEVIER,  
CONFSCI, CROPB, CROPU, DDFB, DDFU, DGENE,  
DISSABS, DRUGB, DRUGLAUNCH,  
DRUGNL, DRUGU, EMBAL, EMBASE, ...' ENTERED AT  
22:17:29 ON 18 JAN 1998

50 FILES IN THE FILE LIST IN STNINDEX

Enter SET DETAIL ON to see search term postings or to view  
search error messages that display as 0\* with SET DETAIL OFF.

=> s (liver?(3a)librar?(3a)human?)

3 FILE AGRICOLA  
1 FILE AIDSLINE  
1 FILE BIOBUSINESS  
428 FILE BIOSIS  
128 FILE BIOTECHABS  
128 FILE BIOTECHDS  
16 FILE CABA  
80 FILE CANCERLIT  
10 FILES SEARCHED...  
559 FILE CAPLUS  
3 FILE CEABA  
83 FILE CJACS  
357 FILE DGENE  
42 FILE DISSABS  
4 FILE DRUGU  
1 FILE EMBAL  
379 FILE EMBASE  
29 FILES SEARCHED...  
45261 FILE GENBANK  
11 FILE IFIPAT  
16 FILE JICST-EPLUS  
34 FILES SEARCHED...  
286 FILE LIFESCI  
450 FILE MEDLINE  
1 FILE NIOSHTIC  
1 FILE NTIS  
39 FILES SEARCHED...  
1 FILE PHAR  
3 FILE PROMT  
175 FILE SCISEARCH  
34 FILE TOXLINE  
415 FILE TOXLIT  
194 FILE USPATFULL  
11 FILE WPIDS  
11 FILE WPINDEX

31 FILES HAVE ONE OR MORE ANSWERS, 50 FILES  
SEARCHED IN STNINDEX

L1 QUE (LIVER?(3A) LIBRAR?(3A) HUMAN?)

=> s l1 and (progesterone? or steroid? or (aromatic hydrocarbon?) or  
dioxin?)

1 FILE AIDSLINE  
1 FILE BIOBUSINESS  
13 FILE BIOSIS  
4 FILE CANCERLIT  
10 FILES SEARCHED...  
17 FILE CAPLUS  
20 FILE CJACS  
3 FILE DGENE  
3 FILE DISSABS  
1 FILE DRUGU  
16 FILE EMBASE  
29 FILES SEARCHED...  
11 FILE GENBANK  
5 FILE IFIPAT  
2 FILE JICST-EPLUS  
6 FILE LIFESCI  
36 FILES SEARCHED...  
19 FILE MEDLINE  
1 FILE NIOSHTIC  
41 FILES SEARCHED...  
7 FILE SCISEARCH  
5 FILE TOXLINE  
11 FILE TOXLIT  
43 FILE USPATFULL  
49 FILES SEARCHED...

20 FILES HAVE ONE OR MORE ANSWERS, 50 FILES  
SEARCHED IN STNINDEX

L2 QUE L1 AND (PROGESTERONE? OR STEROID? OR  
(AROMATIC HYDROCARBON?) OR  
DIOXIN?)

=> file biobusiness

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	
	ENTRY SESSION
FULL ESTIMATED COST	8.55 8.70

FILE 'BIOBUSINESS' ENTERED AT 22:28:52 ON 18 JAN 1998  
COPYRIGHT (C) 1998 Biological Abstracts, Inc. (BIOSIS)

FILE COVERS 1985 TO 16 JAN 1998 (19980116/ED)

This file contains CAS Registry Numbers for easy and accurate  
substance identification.

=> s l2

6598 LIVER?  
862 LIBRAR?  
73980 HUMAN?  
1 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
1491 PROGESTERONE?  
9578 STEROID?  
3331 "AROMATIC"  
3135 HYDROCARBON?  
914 AROMATIC HYDROCARBON?  
("AROMATIC"(W)HYDROCARBON?)  
1194 DIOXIN?

L3 1 L1 AND (PROGESTERONE? OR STEROID? OR  
(AROMATIC HYDROCARBON  
?) OR DIOXIN?)

=> d

L3 ANSWER 1 OF 1 BIOBUSINESS COPYRIGHT 1998  
BIOSIS  
AN 89:30844 BIOBUSINESS

DN 0206649

TI PRODUCTION OF HUMAN RECOMBINANT  
PROAPOLIPOPROTEIN A-I IN ESCHERICHIA  
COLI: PURIFICATION AND BIOCHEMICAL  
CHARACTERIZATION.

AU MOGUILEVSKY N; ROOBOL C; LORIAU R;  
GUILLAUME J-P; JACOBS P; CRAVADOR

A; HERZOG A; BROUWERS L; SCARSO A; ET AL  
CS INQ. DR. A. BOLLEN, SERVICE DE GENETIQUE  
APPLIQUEE U.L.B., RUE DE

L'INDUSTRIE 24, B-1400 NIVELLES, BELGIUM.

SO DNA (NEW YORK), (1989) VOL.8, NO.6, P.429-436.

FS NONUNIQUE

LA ENGLISH

=> d kwic

L3 ANSWER 1 OF 1 BIOBUSINESS COPYRIGHT 1998  
BIOSIS

AB A human liver cDNA library was used to

isolate a clone coding for apolipoprotein A-I (Apo- A-I). The  
clone

carries the sequence for the prepeptide. . .

CC 04300 LIPIDS & RELATED COMPOUNDS; 04600  
PROTEINS & RELATED

COMPOUNDS; 04700 STEROIDS & RELATED  
COMPOUNDS; 52200

BACTERIOLOGY; 55200 INDUSTRIAL MICROBIOLOGY;  
80300 BIOENGINEERING

=> file medline

COST IN U.S. DOLLARS  
TOTAL

SINCE FILE

ENTRY SESSION

FULL ESTIMATED COST 3.20 11.90

FILE 'MEDLINE' ENTERED AT 22:29:43 ON 18 JAN 1998

FILE LAST UPDATED: 16 JAN 1998 (19980116/UP). FILE  
COVERS 1966 TO DATE.

+QLF/CT SHOWS YOU THE ALLOWABLE QUALIFIERS OF  
A TERM.

MEDLINE IS RESUMING UPDATES. NOTICE WILL BE  
GIVEN ONCE THE RELOAD

IS COMPLETED AND RELOAD DETAILS WILL BE FOUND  
IN HELP RLOAD.

THIS FILE CONTAINS CAS REGISTRY NUMBERS FOR  
EASY AND ACCURATE  
SUBSTANCE IDENTIFICATION.

=> s 12

441912 LIVER?

33565 LIBRAR?

6156329 HUMAN?

450 (LIVER?(3A) LIBRAR?(3A) HUMAN?)

47342 PROGESTERONE?

101651 STEROID?

16763 AROMATIC

21475 HYDROCARBON?

3929 AROMATIC HYDROCARBON?

(AROMATIC(W)HYDROCARBON?)

4360 DIOXIN?

L4 19 L1 AND (PROGESTERONE? OR STEROID? OR  
(AROMATIC HYDROCARBON  
?) OR DIOXIN?)

=> d 1-19

L4 ANSWER 1 OF 19 MEDLINE

AN 96170332 MEDLINE

DN 96170332

TI Human dehydroepiandrosterone sulfotransferase. Purification,  
molecular cloning, and characterization.

AU Falany C N; Comer K A; Dooley T P; Glatt H

CS Department of Pharmacology and Toxicology, University of  
Alabama at

Birmingham 35294, USA.

NC GM38953 (NIGMS)

SO ANNALS OF THE NEW YORK ACADEMY OF SCIENCES,  
(1995 Dec 29) 774 59-72.

Journal code: 5NM. ISSN: 0077-8923.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 199606

L4 ANSWER 2 OF 19 MEDLINE

AN 96102134 MEDLINE

DN 96102134

TI A protein that interacts with members of the nuclear hormone  
receptor family: identification and cDNA cloning.

AU Zeiner M; Gehring U

CS Institut für Biologische Chemie, Universität Heidelberg,  
Germany.

SO PROCEEDINGS OF THE NATIONAL ACADEMY OF  
SCIENCES OF THE UNITED STATES

OF AMERICA, (1995 Dec 5) 92 (25) 11465-9.

Journal code: PV3. ISSN: 0027-8424.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-Z35491

EM 199603

L4 ANSWER 3 OF 19 MEDLINE

AN 95322029 MEDLINE

DN 95322029

TI Structural characterization and expression of the human  
dehydroepiandrosterone sulfotransferase gene.

AU Luu-The V; Dufort I; Paquet N; Reimnitz G; Labrie F

CS Medical Research Council Group in Molecular Endocrinology,  
CHUL

Research Center, Quebec, Canada..

SO DNA AND CELL BIOLOGY, (1995 Jun) 14 (6) 511-8.

Journal code: AF9. ISSN: 1044-5498.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-L36191; GENBANK-L36192; GENBANK-  
L36193; GENBANK-L36194;

GENBANK-L36195; GENBANK-L36196

EM 199510

L4 ANSWER 4 OF 19 MEDLINE

AN 95313084 MEDLINE

DN 95313084

TI Type 1 angiotensin II receptors of adrenal tumors.

AU Nawata H; Takayanagi R; Ohnaka K; Sakai Y; Imasaki K;  
Yanase T;

Ikuyama S; Tanaka S; Ohe K

CS Third Department of Internal Medicine, Faculty of Medicine,  
Kyushu

University, Fukuoka, Japan.

SO STEROIDS, (1995 Jan) 60 (1) 28-34.

Journal code: V10. ISSN: 0039-128X.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-S77410

EM 199509

L4 ANSWER 5 OF 19 MEDLINE

AN 95143538 MEDLINE

DN 95143538

TI Fluorescence in situ hybridization analysis of chromosomal localization of three human cytochrome P450 2C genes (CYP2C8, 2C9, and 2C10) at 10q24.1.

AU Inoue K; Inazawa J; Suzuki Y; Shimada T; Yamazaki H; Guengerich F P; Abe T

CS Osaka Prefectural Institute of Public Health, Japan..

SO JAPANESE JOURNAL OF HUMAN GENETICS, (1994 Sep) 39 (3) 337-43.

Journal code: A8U. ISSN: 0916-8478.

CY Japan

DT Journal; Article; (JOURNAL ARTICLE)

LA English

EM 199505

L4 ANSWER 6 OF 19 MEDLINE

AN 95136867 MEDLINE

DN 95136867

TI Stable expression of a human liver UDP-glucuronosyltransferase (UGT2B15) with activity toward steroid and xenobiotic substrates.

AU Green M D; Oturu E M; Tephly T R

CS Department of Pharmacology, University of Iowa, Iowa City 52242..

NC GM 26221 (NIGMS)

SO DRUG METABOLISM AND DISPOSITION, (1994 Sep-Oct) 22 (5) 799-805.

Journal code: EBR. ISSN: 0090-9556.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-U08854

EM 199505

L4 ANSWER 7 OF 19 MEDLINE

AN 94306559 MEDLINE

DN 94306559

TI Human dehydroepiandrosterone sulfotransferase: molecular cloning of cDNA and genomic DNA.

AU Otterness D M; Weinshilboum R

CS Department of Pharmacology, Mayo Medical School, Rochester, MN 55905..

NC GM 28157 (NIGMS)

GM 35720 (NIGMS)

SO CHEMICO-BIOLOGICAL INTERACTIONS, (1994 Jun) 92 (1-3) 145-59. Ref: 41

Journal code: CYV. ISSN: 0009-2797.

CY Ireland

DT Journal; Article; (JOURNAL ARTICLE)

General Review; (REVIEW)  
(REVIEW, TUTORIAL)

LA English

FS Priority Journals; Cancer Journals

EM 199410

L4 ANSWER 8 OF 19 MEDLINE

AN 93238701 MEDLINE

DN 93238701

TI Isozyme developments in mammalian class-I alcohol dehydrogenase.

cDNA cloning, functional correlations, and lack of evidence for genetic isozymes in rabbit.

AU Hoog J O; Vagelopoulos N; Yip P K; Keung W M; Jornvall H

CS Department of Chemistry I, Karolinska Institutet, Stockholm, Sweden..

SO EUROPEAN JOURNAL OF BIOCHEMISTRY, (1993 Apr 1) 213 (1) 31-8.

Journal code: EMZ. ISSN: 0014-2956.

CY GERMANY: Germany, Federal Republic of

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-X69799; GENBANK-D14867; GENBANK-D14868; GENBANK-X71000;

GENBANK-X57950; GENBANK-L09679; GENBANK-L09680; GENBANK-L09681;

GENBANK-L09682; GENBANK-L09683

EM 199307

L4 ANSWER 9 OF 19 MEDLINE

AN 93143674 MEDLINE

DN 93143674

TI Cloning and expression of human liver dehydroepiandrosterone sulphotransferase.

AU Comer K A; Falany J L; Falany C N

CS Department of Pharmacology, University of Rochester, NY 14642..

NC GM38953 (NIGMS)

SO BIOCHEMICAL JOURNAL, (1993 Jan 1) 289 ( Pt 1) 233-40.

Journal code: 9YO. ISSN: 0264-6021.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-L20000; GENBANK-X70222; GENBANK-D17796; GENBANK-D17797;

GENBANK-D17798; GENBANK-D17799; GENBANK-D17800; GENBANK-D17801;

GENBANK-D17802; GENBANK-D17642

EM 199304

L4 ANSWER 10 OF 19 MEDLINE

AN 93140042 MEDLINE

DN 93140042

TI Complementary deoxyribonucleic acid cloning and expression of a

human liver uridine diphosphate-glucuronosyltransferase glucuronidating carboxylic acid-containing drugs.

AU Jin C; Miners J O; Lillywhite K J; Mackenzie P I

CS Department of Clinical Pharmacology, Flinders Medical Centre, Bedford Park, South Australia..

SO JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS, (1993 Jan) 264 (1) 475-9.

Journal code: JP3. ISSN: 0022-3565.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199304

L4 ANSWER 11 OF 19 MEDLINE

AN 91354248 MEDLINE

DN 91354248

TI Cloning and sequencing of cDNA encoding human sepiapterin



reductase—an enzyme involved in tetrahydrobiopterin biosynthesis.  
AU Ichinose H; Katoh S; Sueoka T; Titani K; Fujita K; Nagatsu T  
CS Department of Biochemistry, Nagoya University School of Medicine,  
Japan..

SO BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, (1991 Aug 30)

179 (1) 183-9.

Journal code: 9Y8. ISSN: 0006-291X.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-M76231; GENBANK-S57567; GENBANK-M81108; GENBANK-M81109;

GENBANK-S57525; GENBANK-M74785; GENBANK-M74786; GENBANK-M74787;

GENBANK-M74788; GENBANK-M74790

EM 199112

L4 ANSWER 12 OF 19 MEDLINE

AN 91259898 MEDLINE

DN 91259898

TI Molecular biology of type A endogenous retrovirus.

AU Ono M

CS Department of Molecular Biology, School of Medicine, Kitasato University, Kanagawa, Japan.

SO KITASATO ARCHIVES OF EXPERIMENTAL MEDICINE, (1990 Sep) 63 (2-3)

77-90. Ref: 74

Journal code: KVS. ISSN: 0023-1924.

CY Japan

DT Journal; Article; (JOURNAL ARTICLE)

General Review; (REVIEW)

(REVIEW, ACADEMIC)

LA English

EM 199109

L4 ANSWER 13 OF 19 MEDLINE

AN 90234709 MEDLINE

DN 90234709

TI Nucleotide and deduced amino acid sequence of a human cDNA (NQO2)

corresponding to a second member of the NAD(P)H:quinone oxidoreductase gene family. Extensive polymorphism at the NQO2 gene

locus on chromosome 6.

AU Jaiswal A K; Burnett P; Adesnik M; McBride O W

CS Department of Cell Biology, New York University Medical Center, New

York 10016.

NC GM30701 (NIGMS)

SO BIOCHEMISTRY, (1990 Feb 20) 29 (7) 1899-906.

Journal code: A0G. ISSN: 0006-2960.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-J02888

EM 199008

L4 ANSWER 14 OF 19 MEDLINE

AN 88065931 MEDLINE

DN 88065931

TI A novel steroid thyroid hormone receptor-related gene

inappropriately expressed in human hepatocellular carcinoma.

AU de The H; Marchio A; Tiollais P; Dejean A

CS Unite de Recombinaison et Expression Genetique (INSERM U.163, CNRS

UA 271), Institut Pasteur, Paris, France..

SO NATURE, (1987 Dec 17-23) 330 (6149) 667-70.

Journal code: NSC. ISSN: 0028-0836.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 198803

L4 ANSWER 15 OF 19 MEDLINE

AN 87276521 MEDLINE

DN 87276521

TI Characterization of a cDNA coding for sex steroid-binding protein of human plasma.

AU Que B G; Petra P H

NC HD13956 (NICHD)

SO FEBS LETTERS, (1987 Jul 27) 219 (2) 405-9.

Journal code: EUH. ISSN: 0014-5793.

CY Netherlands

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-X05792

EM 198711

L4 ANSWER 16 OF 19 MEDLINE

AN 87190990 MEDLINE

DN 87190990

TI The cDNA-deduced primary structure of human sex hormone-binding

globulin and location of its steroid-binding domain.

AU Hammond G L; Underhill D A; Smith C L; Goping I S; Harley M J; Musto

N A; Cheng C Y; Bardin C W

NC HD 13541 (NICHD)

SO FEBS LETTERS, (1987 May 4) 215 (1) 100-4.

Journal code: EUH. ISSN: 0014-5793.

CY Netherlands

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-X05403

EM 198708

L4 ANSWER 17 OF 19 MEDLINE

AN 87185403 MEDLINE

DN 87185403

TI Cloning and sequence determination of a complementary DNA related to

human liver microsomal cytochrome P-450 S-mephenytoin 4-hydroxylase.

AU Umbenhauer D R; Martin M V; Lloyd R S; Guengerich F P

NC CA 30907 (NCI)

ES 00267 (NIEHS)

ES 05340 (NIEHS)

SO BIOCHEMISTRY, (1987 Feb 24) 26 (4) 1094-9.

Journal code: A0G. ISSN: 0006-2960.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-M15331

EM 198708

L4 ANSWER 18 OF 19 MEDLINE

AN 86259780 MEDLINE

DN 86259780

TI Complete cDNA sequence of a cytochrome P-450 inducible by glucocorticoids in human liver.

AU Molowa D T; Schuetz E G; Wrighton S A; Watkins P B; Kremers P;

Mendez-Picon G; Parker G A; Guzelian P S

NC AM 18976 (NIADDK)  
 ES-07087 (NIEHS)  
 SO PROCEEDINGS OF THE NATIONAL ACADEMY OF  
 SCIENCES OF THE UNITED STATES  
 OF AMERICA, (1986 Jul) 83 (14) 5311-5.  
 Journal code: PV3. ISSN: 0027-8424.  
 CY United States  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals; Cancer Journals  
 OS GENBANK-M13785  
 EM 198610

L4 ANSWER 19 OF 19 MEDLINE  
 AN 86081170 MEDLINE  
 DN 86081170  
 TI Cloning and isolation of human cytochrome P-450 cDNAs  
 homologous to  
 dioxin-inducible rabbit mRNAs encoding P-450 4 and P-450 6.  
 AU Quattrochi L C; Okino S T; Pendurthi U R; Tukey R H  
 NC CA37139 (NCI)  
 SO DNA, (1985 Oct) 4 (5) 395-400.  
 Journal code: EAW. ISSN: 0198-0238.  
 CY United States  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-M12078; GENBANK-M12079  
 EM 198604

=> s l1 and progesterone?

441912 LIVER?  
 33565 LIBRAR?  
 6156329 HUMAN?  
 450 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
 47342 PROGESTERONE?

L5 3 L1 AND PROGESTERONE?

=> d 1-3

L5 ANSWER 1 OF 3 MEDLINE  
 AN 96102134 MEDLINE  
 DN 96102134  
 TI A protein that interacts with members of the nuclear hormone  
 receptor family: identification and cDNA cloning.  
 AU Zeiner M; Gehring U  
 CS Institut für Biologische Chemie, Universität Heidelberg,  
 Germany.  
 SO PROCEEDINGS OF THE NATIONAL ACADEMY OF  
 SCIENCES OF THE UNITED STATES  
 OF AMERICA, (1995 Dec 5) 92 (25) 11465-9.  
 Journal code: PV3. ISSN: 0027-8424.  
 CY United States  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals; Cancer Journals  
 OS GENBANK-Z35491  
 EM 199603

L5 ANSWER 2 OF 3 MEDLINE  
 AN 91259898 MEDLINE  
 DN 91259898  
 TI Molecular biology of type A endogenous retrovirus.  
 AU Ono M  
 CS Department of Molecular Biology, School of Medicine, Kitasato  
 University, Kanagawa, Japan.  
 SO KITASATO ARCHIVES OF EXPERIMENTAL MEDICINE,  
 (1990 Sep) 63 (2-3)  
 77-90. Ref: 74

Journal code: KVS. ISSN: 0023-1924.  
 CY Japan  
 DT Journal; Article; (JOURNAL ARTICLE)  
 General Review; (REVIEW)  
 (REVIEW, ACADEMIC)  
 LA English  
 EM 199109

L5 ANSWER 3 OF 3 MEDLINE  
 AN 87185403 MEDLINE  
 DN 87185403  
 TI Cloning and sequence determination of a complementary DNA  
 related to  
 human liver microsomal cytochrome P-450 S-mephenytoin 4-  
 hydroxylase.  
 AU Umbenhauer D R; Martin M V; Lloyd R S; Guengerich F P  
 NC CA 30907 (NCI)  
 ES 00267 (NIEHS)  
 ES 05340 (NIEHS)  
 SO BIOCHEMISTRY, (1987 Feb 24) 26 (4) 1094-9.  
 Journal code: A0G. ISSN: 0006-2960.  
 CY United States  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-M15331  
 EM 198708

=> s l1 and (dioxin or (aromatic hydrocarbon?))

441912 LIVER?  
 33565 LIBRAR?  
 6156329 HUMAN?  
 450 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
 2905 DIOXIN  
 16763 AROMATIC  
 21475 HYDROCARBON?  
 3929 AROMATIC HYDROCARBON?  
 (AROMATIC(W)HYDROCARBON?)

L6 2 L1 AND (DIOXIN OR (AROMATIC  
 HYDROCARBON?))

=> d 1-2

L6 ANSWER 1 OF 2 MEDLINE  
 AN 90234709 MEDLINE  
 DN 90234709  
 TI Nucleotide and deduced amino acid sequence of a human cDNA  
 (NQO2)  
 corresponding to a second member of the NAD(P)H:quinone  
 oxidoreductase gene family. Extensive polymorphism at the  
 NQO2 gene  
 locus on chromosome 6.  
 AU Jaiswal A K; Burnett P; Adesnik M; McBride O W  
 CS Department of Cell Biology, New York University Medical  
 Center, New  
 York 10016.  
 NC GM30701 (NIGMS)  
 SO BIOCHEMISTRY, (1990 Feb 20) 29 (7) 1899-906.  
 Journal code: A0G. ISSN: 0006-2960.  
 CY United States  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-J02888  
 EM 199008

L6 ANSWER 2 OF 2 MEDLINE  
 AN 86081170 MEDLINE  
 DN 86081170

TI Cloning and isolation of human cytochrome P-450 cDNAs homologous to dioxin-inducible rabbit mRNAs encoding P-450 4 and P-450 6.  
AU Quattrochi L C; Okino S T; Pendurthi U R; Tukey R H  
NC CA37139 (NCI)  
SO DNA, (1985 Oct) 4 (5) 395-400.  
Journal code: EAW. ISSN: 0198-0238.  
CY United States  
DT Journal; Article; (JOURNAL ARTICLE)  
LA English  
FS Priority Journals  
OS GENBANK-M12078; GENBANK-M12079  
EM 198604

=> file uspatfull

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	ENTRY SESSION
FULL ESTIMATED COST	4.74 16.64

FILE 'USPATFULL' ENTERED AT 22:34:12 ON 18 JAN 1998  
CA INDEXING COPYRIGHT (C) 1998 AMERICAN  
CHEMICAL SOCIETY (ACS)

FILE COVERS 1971 TO PATENT PUBLICATION DATE: 13 Jan 1998 (19980113/PD)  
FILE LAST UPDATED: 14 Jan 1998 (19980114/ED)  
HIGHEST PATENT NUMBER: US5708975  
CA INDEXING IS CURRENT THROUGH 14 Jan 1998 (19980114/UPCA)  
ISSUE CLASS FIELDS (/INCL) CURRENT THROUGH: 13 Jan 1998 (19980113/PD)  
REVISED CLASS FIELDS (/NCL) CURRENT THROUGH: AUG 1997  
USPTO MANUAL OF CLASSIFICATIONS THESAURUS ISSUE DATE: JUN 1997

>>> Page images are available for patents from 1/1/94. Current  
<<<  
>>> week patent text is typically loaded by Thursday morning and  
<<<  
>>> page images are available for display by the end of the day.  
<<<  
>>> Image data for the /FA field are available the following week.  
<<<

>>> Complete CA file indexing for chemical patents (or equivalents)  
<<<  
>>> is included in file records. A thesaurus is available for the <<<  
>>> USPTO Manual of Classifications in the /NCL, /INCL, and  
/RPCL <<<  
>>> fields. This thesaurus includes catchword terms from the  
<<<  
>>> USPTO/MOC subject headings and subheadings. Thesauri are  
also <<<  
>>> available for the WIPO International Patent Classification  
<<<  
>>> (IPC) Manuals, editions 1-6, in the /IC1, /IC2, /IC3, /IC4,  
<<<  
>>> /IC5, and /IC (IC6) fields, respectively. The thesauri in <<<  
>>> the /IC5 and /IC fields include the corresponding catchword  
<<<  
>>> terms from the IPC subject headings and subheadings.  
<<<

This file contains CAS Registry Numbers for easy and accurate  
substance identification.

=> s 16

20473 LIVER?  
15653 LIBRAR?  
171315 HUMAN?  
194 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
861 DIOXIN  
167730 AROMATIC  
197349 HYDROCARBON?  
51324 AROMATIC HYDROCARBON?  
(AROMATIC(W)HYDROCARBON?)  
L7 3 LI AND (DIOXIN OR (AROMATIC  
HYDROCARBON?))

=> d 1-3

L7 ANSWER 1 OF 3 USPATFULL  
AN 96:108836 USPATFULL  
TI Recombinant co-expression system of protein disulfide  
isomerase  
gene, yeast receptor protein ERD2 gene and a foreign product  
polypeptide gene, and a process for producing the foreign  
polypeptide using such system  
IN Hayano, Toshiya, Iruma-gun, Japan  
Katoh, Setsuko, Iruma-gun, Japan  
Takahashi, Nobuhiro, Iruma-gun, Japan  
Suzuki, Masanori, Iruma-gun, Japan  
Honma, Keiichi, Iruma-gun, Japan  
PA Tonen Corporation, Tokyo, Japan (non-U.S. corporation)  
PI US 5578466 961126  
AI US 92-872673 920417 (7)  
PRAI JP 91-114074 910418  
JP 91-311601 911030  
DT Utility  
LN.CNT 2562  
INCL INCLM: 435/069.700  
INCLS: 435/069.100; 435/069.600; 435/254.200  
NCL NCLM: 435/069.700  
NCLS: 435/069.100; 435/069.600; 435/254.200  
IC [6]  
ICM: C12N001-19  
ICS: C12N015-14; C12N015-62  
EXF 435/69.1; 435/254.2; 435/69.6; 435/69.7; 536/23.5;  
536/23.2;  
536/23.4  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L7 ANSWER 2 OF 3 USPATFULL  
AN 95:60284 USPATFULL  
TI Human cell line stably expressing 5cDNAs encoding  
procarcinogen-activating enzymes and related mutagenicity  
assays  
IN Crespi, Charles L., Marblehead, MA, United States  
Penman, Bruce W., Salem, MA, United States  
Davies, Robin L., Amherst, VA, United States  
PA Gentest Corporation, Woburn, MA, United States (U.S.  
corporation)  
PI US 5429948 950704  
AI US 92-997455 921228 (7)  
RLI Continuation-in-part of Ser. No. US 90-597815, filed on 15  
Oct  
1990, now abandoned And Ser. No. US 91-771520, filed on 4  
Oct  
1991, now abandoned which is a continuation of Ser. No. US  
88-162885, filed on 2 Mar 1988, now abandoned, said Ser. No.  
US  
-597815 which is a continuation-in-part of Ser. No. US -  
162885  
DT Utility  
LN.CNT 1828  
INCL INCLM: 435/240.200  
INCLS: 435/172.100; 435/172.300

NCL NCLM: 435/372.000  
NCLS: 435/172.100; 435/172.300  
IC [6]  
ICM: C12N015-00  
EXF 435/240.2  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L7 ANSWER 3 OF 3 USPATFULL  
AN 95:47624 USPATFULL  
TI Methods and compositions for the expression of biologically active fusion proteins comprising a eukaryotic cytochrome P450 fused to a reductase in bacteria  
IN Fisher, Charles W., Dallas, TX, United States  
Barnes, Henry J., Chula Vista, CA, United States  
Estabrook, Ronald W., Dallas, TX, United States  
PA Board of Regents, The University of Texas System, Austin, TX, United States (U.S. corporation)  
PI US 5420027 950530  
AI US 92-908317 920702 (7)  
RLI Continuation of Ser. No. US 91-640473, filed on 10 Jan 1991, now patented, Pat. No. US 5240831  
DT Utility  
LN.CNT 2930  
INCL INCLM: 435/189.000  
INCLS: 435/069.700; 435/252.300; 435/252.330; 435/320.100; 536/023.400; 536/023.200; 935/010.000; 935/014.000; 935/027.000; 935/038.000; 935/044.000; 935/047.000  
NCL NCLM: 435/189.000  
NCLS: 435/069.700; 435/252.300; 435/252.330; 435/320.100; 536/023.200; 536/023.400; 935/010.000; 935/014.000; 935/027.000; 935/038.000; 935/044.000; 935/047.000  
IC [6]  
ICM: C12N009-02  
ICS: C12N015-53; C12N015-62; C12N015-63  
EXF 435/69.1; 435/69.7; 435/189; 435/252.3; 435/252.33; 435/320.1; 536/23.2; 536/23.4  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

=> file biosis

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	
	ENTRY SESSION
FULL ESTIMATED COST	4.05 20.69

FILE 'BIOSIS' ENTERED AT 22:35:11 ON 18 JAN 1998  
COPYRIGHT (C) 1998 BIOSIS(R)

FILE COVERS 1969 TO DATE.  
CAS REGISTRY NUMBERS AND CHEMICAL NAMES (CNs) PRESENT  
FROM JANUARY 1969 TO DATE.

RECORDS LAST ADDED: 14 January 1998 (980114/ED)  
CAS REGISTRY NUMBERS (R) LAST ADDED: 14 January 1998 (980114/UP)

=> s 16

351026 LIVER?  
27005 LIBRAR?  
3986104 HUMAN?  
428 (LIVER?(3A) LIBRAR?(3A) HUMAN?)

5216 DIOXIN  
31727 AROMATIC  
24157 HYDROCARBON?  
8535 AROMATIC HYDROCARBON?  
(AROMATIC(W)HYDROCARBON?)  
L8 3 L1 AND (DIOXIN OR (AROMATIC HYDROCARBON?))

=> d 1-3

L8 ANSWER 1 OF 3 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 90:176099 BIOSIS  
DN BA89:93269  
TI NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCE OF A HUMAN COMPLEMENTARY DNA NQO-2 CORRESPONDING TO A SECOND MEMBER OF THE NADPH QUINONE OXIDOREDUCTASE GENE FAMILY EXTENSIVE POLYMORPHISM AT THE NQO-2 GENE LOCUS ON CHROMOSOME 6.  
AU JAISWAL A K; BURNETT P; ADESNIK M; MCBRIDE O W  
CS DEP. CELL BIOL., NEW YORK UNIV. MED. CENT., 550 FIRST AVE., NEW YORK, N.Y. 10016.  
SO BIOCHEMISTRY 29 (7). 1990. 1899-1906. CODEN: BICHAW ISSN: 0006-2960  
LA English

L8 ANSWER 2 OF 3 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 89:28147 BIOSIS  
DN BA87:16147  
TI HUMAN DIOXIN-INDUCIBLE CYTOSOLIC NADPH MENADIONE OXIDOREDUCTASE COMPLEMENTARY DNA SEQUENCE AND LOCALIZATION OF GENE TO CHROMOSOME 16.  
AU JAISWAL A K; MCBRIDE O W; ADESNIK M; NEBERT D W  
CS DEP. CELL BIOL. AND KAPLAN CANCER CENT., NEW YORK UNIV. MED. CENT., NEW YORK, 10016.  
SO J BIOL CHEM 263 (27). 1988. 13572-13578. CODEN: JBCHA3 ISSN: 0021-9258  
LA English

L8 ANSWER 3 OF 3 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 86:142185 BIOSIS  
DN BA81:52601  
TI CLONING AND ISOLATION OF HUMAN CYTOCHROME P-450 COMPLEMENTARY DNA SPECIES HOMOLOGOUS TO DIOXIN-INDUCIBLE RABBIT MESSENGER RNA SPECIES ENCODING P-450-4 AND P-450-6.  
AU QUATTROCHI L C; OKINO S T; PENDURTHI U R; TUKEY R H  
CS UCSD CANCER CENTER, T-012 UCSD, LA JOLLA, CALIF. 92093.  
SO DNA (N Y) 4 (5). 1985. 395-400. CODEN: DNAADR ISSN: 0198-0238  
LA English

=> s 15

351026 LIVER?  
27005 LIBRAR?  
3986104 HUMAN?  
428 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
53680 PROGESTERONE?  
L9 1 L1 AND PROGESTERONE?

=> d

L9 ANSWER 1 OF 1 BIOSIS COPYRIGHT 1998 BIOSIS  
AN 87:187108 BIOSIS  
DN BA83:95232  
TI CLONING AND SEQUENCE DETERMINATION OF A  
COMPLEMENTARY DNA RELATED TO  
HUMAN LIVER MICROSOMAL CYTOCHROME P-450 5  
MEPHENYTOIN 4-HYDROXYLASE.  
AU UMBENHAUER D R; MARTIN M V; LLOYD R S;  
GUENGERICH F P  
CS DEP. BIOCHEMISTRY, CENT. MOL. TOXICOL.,  
VANDERBILT UNIV. SCH. MED.,  
NASHVILLE, TENN. 37232.  
SO BIOCHEMISTRY 26 (4). 1987. 1094-1099. CODEN:  
BICHAW ISSN: 0006-2960  
LA English

=> uspatfull

USPATFULL IS NOT A RECOGNIZED COMMAND  
The previous command name entered was not recognized by the  
system.

For a list of commands available to you in the current file, enter  
"HELP COMMANDS" at an arrow prompt (=>).

=> file uspatfull

COST IN U.S. DOLLARS	SINCE FILE
TOTAL	ENTRY SESSION
FULL ESTIMATED COST	5.92 26.61

FILE 'USPATFULL' ENTERED AT 22:36:33 ON 18 JAN 1998  
CA INDEXING COPYRIGHT (C) 1998 AMERICAN  
CHEMICAL SOCIETY (ACS)

FILE COVERS 1971 TO PATENT PUBLICATION DATE: 13 Jan  
1998 (19980113/PD)  
FILE LAST UPDATED: 14 Jan 1998 (19980114/ED)  
HIGHEST PATENT NUMBER: US5708975  
CA INDEXING IS CURRENT THROUGH 14 Jan 1998  
(19980114/UPCA)  
ISSUE CLASS FIELDS (/INCL) CURRENT THROUGH: 13 Jan  
1998 (19980113/PD)  
REVISED CLASS FIELDS (/NCL) CURRENT THROUGH: AUG  
1997  
USPTO MANUAL OF CLASSIFICATIONS THESAURUS ISSUE  
DATE: JUN 1997

>>> Page images are available for patents from 1/1/94. Current  
<<<  
>>> week patent text is typically loaded by Thursday morning and  
<<<  
>>> page images are available for display by the end of the day.  
<<<  
>>> Image data for the /FA field are available the following week.  
<<<

>>> Complete CA file indexing for chemical patents (or equivalents)  
<<<  
>>> is included in file records. A thesaurus is available for the <<<  
>>> USPTO Manual of Classifications in the /NCL, /INCL, and  
/RPCL <<<  
>>> fields. This thesaurus includes catchword terms from the  
<<<  
>>> USPTO/MOC subject headings and subheadings. Thesauri are  
also <<<

>>> available for the WIPO International Patent Classification  
<<<  
>>> (IPC) Manuals, editions 1-6, in the /IC1, /IC2, /IC3, /IC4,  
<<<  
>>> /IC5, and /IC (/IC6) fields, respectively. The thesauri in <<<  
>>> the /IC5 and /IC fields include the corresponding catchword  
<<<  
>>> terms from the IPC subject headings and subheadings.  
<<<

This file contains CAS Registry Numbers for easy and accurate  
substance identification.

=> s 15

20473 LIVER?  
15653 LIBRAR?  
171315 HUMAN?  
194 (LIVER?(3A) LIBRAR?(3A) HUMAN?)  
2684 PROGESTERONE?

L10 11 L1 AND PROGESTERONE?

=> d 1-11

L10 ANSWER 1 OF 11 USPATFULL  
AN 97:115390 USPATFULL  
TI Orphan steroid hormone receptors  
IN Evans, Ronald M., La Jolla, CA, United States  
Mangelsdorf, David J., San Diego, CA, United States  
Ong, Estelita S., San Diego, CA, United States  
Oro, Anthony E., San Diego, CA, United States  
Borgmeyer, Uwe K., Hamburg, Germany, Federal Republic of  
Giguere, Vincent, Etobicoke, Canada  
Yao, Tso-Pang, San Diego, CA, United States  
PA The Salk Institute for Biological Studies, La Jolla, CA, United  
States (U.S. corporation)  
PI US 5696233 971209  
AI US 95-463694 950605 (8)  
RLI Division of Ser. No. US 94-333358, filed on 2 Nov 1994, now  
patented, Pat. No. US 5571696 which is a continuation of Ser.  
No.  
US 91-761068, filed on 17 Sep 1991, now abandoned  
DT Utility  
LN.CNT 1535  
INCL INCLM: 530/350.000  
INCL: 530/358.000  
NCL NCLM: 530/350.000  
NCL: 530/358.000  
IC [6]  
ICM: C07K014-72  
EXF 530/350; 530/399; 530/358

L10 ANSWER 2 OF 11 USPATFULL  
AN 97:104598 USPATFULL  
TI Constitutive activator of retinoid acid response (car) receptor  
fusion protien  
IN Moore, David D., Hingham, MA, United States  
Baes, Myriam I., Belmont, MA, United States  
PA The General Hospital Corporation, Boston, MA, United States  
(U.S.  
corporation)  
PI US 5686574 971111  
AI US 95-459489 950605 (8)  
RLI Division of Ser. No. US 92-843350, filed on 26 Feb 1992  
DT Utility  
LN.CNT 969  
INCL INCLM: 530/350.000  
INCL: 536/023.400; 435/069.100; 435/069.700  
NCL NCLM: 530/350.000  
NCL: 435/069.100; 435/069.700; 536/023.400

IC [6]  
 ICM: C07K014-435  
 ICS: C12N015-62; C12N015-00; C12P021-00  
 EXF 530/350; 514/2; 435/69.1; 435/69.7; 536/23.4  
 CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 3 OF 11 USPATFULL  
 AN 97:5949 USPATFULL  
 TI Variants of human corticosteroid binding globulin  
 IN Hammond, Geoffrey L., Lambeth, Canada  
 Avvakumov, George V., London, Canada  
 PA Allelix Biopharmaceutical Inc., Ontario, Canada (non-U.S. corporation)  
 PI US 5595969 970121  
 AI US 95-421891 950414 (8)  
 RLI Division of Ser. No. US 92-994423, filed on 16 Dec 1992, now patented, Pat. No. US 5432080  
 DT Utility  
 LN.CNT 905  
 INCL INCLM: 514/008.000  
 INCLS: 514/021.000; 530/386.000; 530/395.000  
 NCL NCLM: 514/008.000  
 NCLS: 514/021.000; 530/386.000; 530/395.000  
 IC [6]  
 ICM: A61K037-04  
 EXF 530/386; 530/395; 514/8; 514/21; 935/10  
 CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 4 OF 11 USPATFULL  
 AN 96:101464 USPATFULL  
 TI Receptors  
 IN Evans, Ronald M., La Jolla, CA, United States  
 Mangelsdorf, David J., San Diego, CA, United States  
 Ong, Estelita S., San Diego, CA, United States  
 Oro, Anthony E., San Diego, CA, United States  
 Borgmeyer, Uwe K., Hamburg, Germany, Federal Republic of  
 Giguere, Vincent, Etobicoke, Canada  
 Yao, Tso-Pang, San Diego, CA, United States  
 PA The Salk Institute for Biological Studies, La Jolla, CA, United States (U.S. corporation)  
 PI US 5571696 961105  
 AI US 94-333358 941102 (8)  
 RLI Continuation of Ser. No. US 91-761068, filed on 17 Sep 1991, now abandoned  
 DT Utility  
 LN.CNT 1555  
 INCL INCLM: 435/069.100  
 INCLS: 435/240.100; 435/320.100; 536/023.100; 536/023.400  
 NCL NCLM: 435/069.100  
 NCLS: 435/320.100; 435/325.000; 536/023.100; 536/023.400  
 IC [6]  
 ICM: C12P021-06  
 ICS: C12N005-00; C12N015-00; C07H017-00  
 EXF 536/23.1; 536/23.4; 435/69.1; 435/69.7; 435/240.1; 435/320.1  
 CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 5 OF 11 USPATFULL  
 AN 95:103380 USPATFULL  
 TI Steroid/thyroid hormone receptor-related gene, which is inappropriately expressed in human hepatocellular carcinoma, and which is a retinoic acid receptor  
 IN Blandin De The, Hughes, Faculty of Medicine, 75003 Paris, France  
 Marchio, Agnes, Faculty of Medicine, 75011 Paris, France  
 Tiollais, Pierre, Faculty of Medicine, 75013 Paris, France  
 DeJean, Anne, Faculty of Medicine, 75014 Paris, France  
 Brand, Nigel, Faculty of Medicine, 67085 Strasbourg, France

Petkovich, Martin, Faculty of Medicine, 67085 Strasbourg, France  
 Krust, Andree, Faculty of Medicine, 67085 Strasbourg, France  
 Chambon, Pierre, Faculty of Medicine, 67085 Strasbourg, France

PI US 5468617 951121  
 AI US 94-190555 940202 (8)  
 RLI Division of Ser. No. US 93-95706, filed on 22 Jul 1993, now patented, Pat. No. US 5358848 which is a division of Ser. No. US 92-989902, filed on 11 Dec 1992, now patented, Pat. No. US 5317090 which is a continuation of Ser. No. US 92-860577, filed on 30 Mar 1992, now abandoned which is a continuation of Ser. No. US 91-751612, filed on 21 Aug 1991, now abandoned which is a continuation of Ser. No. US 89-330405, filed on 30 Mar 1989, now abandoned which is a continuation-in-part of Ser. No. US 88-278136, filed on 30 Nov 1988, now abandoned which is a continuation-in-part of Ser. No. US 88-209009, filed on 20 Jun 1988, now patented, Pat. No. US 5149781 which is a continuation-in-part of Ser. No. US 87-134130, filed on 17 Dec 1987, now patented, Pat. No. US 5223606 which is a continuation-in-part of Ser. No. US 87-133687, filed on 16 Dec 1987, now abandoned  
 DT Utility  
 LN.CNT 2011  
 INCL INCLM: 435/007.800  
 INCLS: 435/007.100; 530/350.000; 436/063.000; 436/501.000  
 NCL NCLM: 435/007.800  
 NCLS: 435/007.100; 436/063.000; 436/501.000; 530/350.000  
 IC [6]  
 ICM: G01N033-48  
 EXF 435/7.1; 435/7.8; 435/172.3; 435/810; 436/63; 436/501; 530/350  
 CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 6 OF 11 USPATFULL  
 AN 95:62633 USPATFULL  
 TI DNA encoding variants of human corticosteroid binding globulin  
 IN Hammond, Geoffrey L., Lambeth, Canada  
 Avvakumov, George V., London, Canada  
 PA Allelix Biopharmaceuticals Inc., Mississauga, Canada (non-U.S. corporation)  
 PI US 5432080 950711  
 AI US 92-994423 921216 (7)  
 DT Utility  
 LN.CNT 896  
 INCL INCLM: 435/252.300  
 INCLS: 536/023.500; 530/386.000  
 NCL NCLM: 435/252.300  
 NCLS: 530/386.000; 536/023.500  
 IC [6]  
 ICM: C12N005-10  
 ICS: C12N015-00  
 EXF 536/23.5; 435/252.3; 530/386  
 CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 7 OF 11 USPATFULL  
 AN 95:47624 USPATFULL  
 TI Methods and compositions for the expression of biologically active fusion proteins comprising a eukaryotic cytochrome P450 fused to a reductase in bacteria  
 IN Fisher, Charles W., Dallas, TX, United States  
 Barnes, Henry J., Chula Vista, CA, United States  
 Estabrook, Ronald W., Dallas, TX, United States

PA Board of Regents, The University of Texas System, Austin,  
TX,  
United States (U.S. corporation)  
PI US 5420027 950530  
AI US 92-908317 920702 (7)  
RLI Continuation of Ser. No. US 91-640473, filed on 10 Jan 1991,  
now

patented, Pat. No. US 5240831  
DT Utility  
LN.CNT 2930  
INCL INCLM: 435/189.000  
INCLS: 435/069.700; 435/252.300; 435/252.330;  
435/320.100;  
536/023.400; 536/023.200; 935/010.000; 935/014.000;  
935/027.000; 935/038.000; 935/044.000; 935/047.000  
NCL NCLM: 435/189.000  
NCLS: 435/069.700; 435/252.300; 435/252.330;  
435/320.100;  
536/023.200; 536/023.400; 935/010.000; 935/014.000;  
935/027.000; 935/038.000; 935/044.000; 935/047.000

IC [6]  
ICM: C12N009-02  
ICS: C12N015-53; C12N015-62; C12N015-63  
EXF 435/69.1; 435/69.7; 435/189; 435/252.3; 435/252.33;  
435/320.1;  
536/23.2; 536/23.4  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 8 OF 11 USPATFULL  
AN 94:112894 USPATFULL  
TI Steroid/thyroid hormone receptor-related gene, which is  
inappropriately expressed in human hepatocellular carcinoma, and  
which is a retinoic acid receptor

IN De The, Hughes B., Paris, France  
Marchio, Agnes, Paris, France  
Tiollais, Pierre, Paris, France  
DeJean, Anne, Paris, France  
Brand, Nigel, Strasbourg, France  
Petkovich, Martin, Strasbourg, France  
Krust, Andree, Strasbourg, France  
Chambon, Pierre, Strasbourg, France

PA Institut Pasteur, Paris Cedex, France (non-U.S. corporation)  
PI US 5376530 941227  
AI US 93-95706 930722 (8)  
RLI Division of Ser. No. US 92-989902, filed on 11 Dec 1992  
which is a  
continuation of Ser. No. US 92-860577, filed on 30 Mar 1992,  
now  
abandoned which is a continuation of Ser. No. US 91-751612,  
filed  
on 21 Aug 1991, now abandoned which is a continuation of Ser.  
No.

US 89-330405, filed on 30 Mar 1989, now abandoned which is  
a  
continuation-in-part of Ser. No. US 88-278136, filed on 30 Nov  
1988, now abandoned which is a continuation-in-part of Ser. No.

US  
88-209009, filed on 20 Jun 1988, now patented, Pat. No. US  
5149781

which is a continuation-in-part of Ser. No. US 87-134130, filed  
on  
17 Dec 1987, now patented, Pat. No. US 5223606 which is a  
continuation-in-part of Ser. No. US 87-133687, filed on 16 Dec  
1987, now abandoned

DT Utility  
LN.CNT 1943  
INCL INCLM: 435/006.000  
INCLS: 435/069.100; 435/172.300; 536/023.100;  
536/024.310;  
530/326.000; 530/327.000; 530/350.000; 530/828.000  
NCL NCLM: 435/006.000

NCLS: 435/069.100; 435/172.300; 530/326.000;  
530/327.000;  
530/350.000; 530/828.000; 536/023.100; 536/024.310  
IC [5]  
ICM: C12Q001-68  
EXF 435/6; 435/91; 435/69.1; 435/172.3; 536/23.1; 536/24.31;  
530/326;  
530/327; 530/350; 530/828; 530/846  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 9 OF 11 USPATFULL  
AN 94:47046 USPATFULL  
TI Steroid/thyroid hormone receptor-related gene, which is  
inappropriately expressed in human hepatocellular carcinoma,  
and

which is a retinoic acid receptor  
IN Blaudin De The, Hughes, Paris, France  
Marchio, Agnes, Paris, France  
Tiollais, Pierre, Paris, France  
DeJean, Anne, Paris, France  
Brand, Nigel, Strasbourg, France  
Petkovich, Martin, Strasbourg, France  
Krust, Andree, Strasbourg, France  
Chambon, Pierre, Strasbourg, France

PA Institut Pasteur, Paris, France (non-U.S. corporation)  
PI US 5317090 940531  
AI US 92-989902 921211 (7)  
RLI Continuation of Ser. No. US 92-860577, filed on 30 Mar  
1992, now  
abandoned which is a continuation of Ser. No. US 91-751612,  
filed  
on 21 Aug 1991, now abandoned which is a continuation of Ser.  
No.

US 89-330405, filed on 30 Mar 1989, now abandoned which is  
a  
continuation-in-part of Ser. No. US 88-278136, filed on 30 Nov  
1988, now abandoned which is a continuation-in-part of Ser. No.

US  
88-209009, filed on 20 Jun 1988, now patented, Pat. No. US  
5149781  
which is a continuation-in-part of Ser. No. US 87-134130, filed  
on

17 Dec 1987 And Ser. No. US 87-133687, filed on 16 Dec  
1987, now  
abandoned

DT Utility  
LN.CNT 1892  
INCL INCLM: 530/387.100  
INCLS: 530/387.900; 530/388.100; 530/388.220; 530/391.100  
NCL NCLM: 530/387.100  
NCLS: 530/387.900; 530/388.100; 530/388.220; 530/391.100  
IC [5]

ICM: A61K035-14  
EXF 530/387.1; 530/387.9; 530/388.1; 530/388.2; 530/388.22;  
530/391.1;  
530/864; 435/7.1; 435/7.92; 435/70.2; 435/70.21; 435/172.2;  
436/501; 436/547; 436/548  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 10 OF 11 USPATFULL  
AN 93:52684 USPATFULL  
TI Steroid/thyroid hormone receptor-related protein  
inappropriately

expressed in human hepatocellular carcinoma  
IN Blaudin de The, Hughes, Paris, France  
Marchio, Agnes, Paris, France  
Tiollais, Pierre, Paris, France  
DeJean, Anne, Paris, France

PA Institut Pasteur, Paris, France (non-U.S. corporation)  
PI US 5223606 930629  
AI US 87-134130 871217 (7)

RLI Continuation-in-part of Ser. No. US 87-133687, filed on 16  
Dec

1987, now abandoned

DT Utility

LN.CNT 862

INCL INCLM: 530/350.000

INCLS: 530/828.000; 530/846.000; 435/069.100

NCL NCLM: 530/350.000

NCLS: 435/069.100; 530/828.000; 530/846.000

IC [5]

ICM: C07K013-00

EXF 530/350; 530/828; 530/846

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

L10 ANSWER 11 OF 11 USPATFULL

AN 92:78970 USPATFULL

TI Steroid/thyroid hormone receptor-related gene inappropriately  
expressed in human hepatocellular carcinoma

IN Blaudin de THE, Hugues, Paris, France

Marchio, Agnes, Paris, France

Tiollais, Pierre, Paris, France

Dejean, Anne, Paris, France

PA Institut Pasteur, United States (non-U.S. corporation)

PI US 5149781 920922

AI US 88-209009 880620 (7)

RLI Continuation-in-part of Ser. No. US 87-134130, filed on 17

Dec

1987 which is a continuation-in-part of Ser. No. US 87-133687,  
filed on 16 Dec 1987, now abandoned

DT Utility

LN.CNT 986

INCL INCLM: 530/326.000

INCLS: 530/327.000; 530/328.000; 530/350.000; 530/405.000

NCL NCLM: 530/326.000

NCLS: 530/327.000; 530/328.000; 530/350.000; 530/405.000

IC [5]

ICM: C07K007-06

ICS: C07K007-08; C07K007-10; C07K017-00; C07K013-00

EXF 530/350; 530/327; 530/328; 530/326; 530/405

CAS INDEXING IS AVAILABLE FOR THIS PATENT.



FILE 'MEDLINE' ENTERED AT 11:23:02 ON 19 JAN 1998

FILE LAST UPDATED: 16 JAN 1998 (19980116/UP). FILE COVERS 1966 TO DATE.  
+QLF/CT SHOWS YOU THE ALLOWABLE QUALIFIERS OF A TERM.

MEDLINE IS RESUMING UPDATES. NOTICE WILL BE GIVEN ONCE THE RELOAD  
IS COMPLETED AND RELOAD DETAILS WILL BE FOUND IN HELP RLOAD.

THIS FILE CONTAINS CAS REGISTRY NUMBERS FOR EASY AND ACCURATE  
SUBSTANCE IDENTIFICATION.

=> e purnelle b/au

E1	9	PURNELL W D/AU
E2	1	PURNELL W L/AU
E3	22 -->	PURNELLE B/AU
E4	3	PURNELLE G/AU
E5	4	PURNER F/AU
E6	1	PURNER J/AU
E7	5	PURNER M B/AU
E8	1	PURNEY J W JR/AU
E9	1	PURNEY R G/AU
E10	2	PURNICHESCU M/AU
E11	4	PURNIMA/AU
E12	2	PURNIMA G/AU

=> s e3

L1 22 "PURNELLE B"/AU

=> e goffeau a/au

E1	2	GOFFE T/AU
E2	2	GOFFE T R/AU
E3	181 -->	GOFFEAU A/AU
E4	6	GOFFEN B S/AU
E5	1	GOFFEN S V/AU
E6	2	GOFFENBERG S I/AU
E7	1	GOFFENEY B/AU
E8	1	GOFFENEY N W/AU
E9	3	GOFFENG A R/AU
E10	2	GOFFENG L/AU
E11	2	GOFFENG L O/AU
E12	1	GOFFER D/AU

=> S LE3

L2 7 LE3

=> DEL L2

DELETE L2? (Y)/N:Y

=> S E3

L2 181 "GOFFEAU A"/AU

=> S L1 AND L2

=&gt; D 1-20

L3 ANSWER 1 OF 20 MEDLINE

AN 97197984 MEDLINE

DN 97197984

TI The sequence of 32b on the left arm of yeast chromosome XII reveals six known genes, a new member of the seripauperins family and a new ABS transporter homologous to the human multidrug resistance protein.

AU **Purnelle B; Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

SO YEAST, (1997 Feb) 13 (2) 183-8.

Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-X97560

EM 199707

EW 19970703

L3 ANSWER 2 OF 20 MEDLINE

AN 97197982 MEDLINE

DN 97197982

TI An 18.3 kb DNA fragment from yeast chromosome VII carries four unknown open reading frames, the gene for an Asn synthase, remnants of Ty and three tRNA genes.

AU Van Dyck L; Tettelin H; **Purnelle B; Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

SO YEAST, (1997 Feb) 13 (2) 171-6.

Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-X83099; SWISSPROT-P22106; SWISSPROT-P31752; SWISSPROT-P08243

EM 199707

EW 19970703

L3 ANSWER 3 OF 20 MEDLINE

AN 97103777 MEDLINE

DN 97103777

TI The sequence of 55 kb on the left arm of yeast chromosome XVI identifies a small nuclear RNA, a new putative protein kinase and two new putative regulators.

AU **Purnelle B; Coster F; Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

SO YEAST, (1996 Nov) 12 (14) 1483-92.

Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-X96770

EM 199705

EW 19970504

L3 ANSWER 4 OF 20 MEDLINE

AN 97103776 MEDLINE  
 DN 97103776  
 TI Nucleotide sequence analysis of a 40 kb segment on the right arm of yeast chromosome XV reveals 18 open reading frames including a new pyruvate kinase and three homologues to chromosome I genes.  
 AU **Purnelle B; Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.  
 SO YEAST, (1996 Nov) 12 (14) 1475-81.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X95720  
 EM 199705  
 EW 19970504

L3 ANSWER 5 OF 20 MEDLINE  
 AN 96021610 MEDLINE  
 DN 96021610  
 TI An 8.2 kb DNA segment from chromosome XIV carries the RPD3 and PAS8 genes as well as the Saccharomyces cerevisiae homologue of the thiamine-repressed nmt1 gene and a chromosome III-duplicated gene for a putative aryl-alcohol dehydrogenase.  
 AU Van Dyck L; Pascual-Ahuir A; **Purnelle B; Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Belgium.  
 SO YEAST, (1995 Aug) 11 (10) 987-91.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X83226; GENBANK-Z46259; GENBANK-L08964  
 EM 199604

L3 ANSWER 6 OF 20 MEDLINE  
 AN 95282517 MEDLINE  
 DN 95282517  
 TI The sequence of a 13.5 kb DNA segment from the left arm of yeast chromosome XIV reveals MER1; RAP1; a new putative member of the DNA replication complex and a new putative serine/threonine phosphatase gene.  
 AU Coster F; Van Dyck L; Jonniaux J L; **Purnelle B; Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.  
 SO YEAST, (1995 Jan) 11 (1) 85-91.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X78898  
 EM 199509

L3 ANSWER 7 OF 20 MEDLINE  
 AN 95274326 MEDLINE  
 DN 95274326  
 TI The sequence of a 36 kb segment on the left arm of yeast chromosome X identifies 24 open reading frames including NUC1, PRP21 (SPP91), CDC6, CRY2, the gene for S24, a homologue to the aconitase gene ACO1 and two homologues to chromosome III genes.  
 AU **Purnelle B; Coster F; Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain,  
 Belgium..  
 SO YEAST, (1994 Sep) 10 (9) 1235-49.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X77688  
 EM 199508

L3 ANSWER 8 OF 20 MEDLINE  
 AN 95242839 MEDLINE  
 DN 95242839  
 TI A 21.7 kb DNA segment on the left arm of yeast chromosome XIV  
 carries WHI3, GCR2, SPX18, SPX19, an homologue to the heat shock  
 gene SSB1 and 8 new open reading frames of unknown function.  
 AU Jonniaux J L; Coster F; **Purnelle B**; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain,  
 Louvain-la-Neuve, Belgium..  
 SO YEAST, (1994 Dec) 10 (12) 1639-45.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X78898  
 EM 199507

L3 ANSWER 9 OF 20 MEDLINE  
 AN 95176707 MEDLINE  
 DN 95176707  
 TI The sequence of a 22.4 kb DNA fragment from the left arm of yeast  
 chromosome II reveals homologues to bacterial proline synthetase and  
 murine alpha-adaptin, as well as a new permease and a DNA-binding  
 protein.  
 AU De Wergifosse P; Jacques B; Jonniaux J L; **Purnelle B**;  
 Skala J; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain,  
 Belgium..  
 SO YEAST, (1994 Nov) 10 (11) 1489-96.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-X78214; GENBANK-X78217  
 EM 199506

L3 ANSWER 10 OF 20 MEDLINE  
 AN 94378718 MEDLINE  
 DN 94378718  
 TI The sequence of an 8.8 kb segment on the left arm of chromosome II  
 from Saccharomyces cerevisiae reveals four new open reading frames  
 including homologs of animal DNA polymerase alpha-primases and  
 bacterial GTP cyclohydrolase II.  
 AU Skala J; Van Dyck L; **Purnelle B**; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain,  
 Louvain-la-Neuve, Belgium..  
 SO YEAST, (1994 Apr) 10 Suppl A S13-24.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals

OS GENBANK-X74738; GENBANK-S58377  
EM 199412

L3 ANSWER 11 OF 20 MEDLINE

AN 94262309 MEDLINE

DN 94262309

TI Analysis of an 11.7 kb DNA fragment of chromosome XI reveals a new tRNA gene and four new open reading frames including a leucine zipper protein and a homologue to the yeast mitochondrial regulator ABF2.

AU **Purnelle B**; Skala J; van Dyck L; **Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

SO YEAST, (1994 Jan) 10 (1) 125-30.  
Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-X71622

EM 199409

L3 ANSWER 12 OF 20 MEDLINE

AN 94205268 MEDLINE

DN 94205268

TI The sequence of a 17.5 kb DNA fragment on the left arm of yeast chromosome XI identifies the protein kinase gene ELM1, the DNA primase gene PRI2, a new gene encoding a putative histone and seven new open reading frames.

AU **Purnelle B**; Tettelin H; Van Dyck L; Skala J; **Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium..

SO YEAST, (1993 Dec) 9 (12) 1379-84.  
Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-X71621

EM 199407

L3 ANSWER 13 OF 20 MEDLINE

AN 93127731 MEDLINE

DN 93127731

TI The sequence of a 12 kb fragment on the left arm of yeast chromosome XI reveals five new open reading frames, including a zinc finger protein and a homolog of the UDP-glucose pyrophosphorylase from potato.

AU **Purnelle B**; Skala J; Van Dyck L; **Goffeau A**

CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

SO YEAST, (1992 Nov) 8 (11) 977-86.  
Journal code: YEA. ISSN: 0749-503X.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199304

L3 ANSWER 14 OF 20 MEDLINE

AN 93070615 MEDLINE

DN 93070615

TI The sequence of an 8 kb segment on the left arm of chromosome II from *Saccharomyces cerevisiae* identifies five new open reading

frames of unknown functions, two tRNA genes and two transposable elements.

AU Skala J; Van Dyck L; **Purnelle B**; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Belgium..  
 SO YEAST, (1992 Sep) 8 (9) 777-85.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 EM 199302

L3 ANSWER 15 OF 20 MEDLINE  
 AN 93070614 MEDLINE  
 DN 93070614  
 TI An 11.4 kb DNA segment on the left arm of yeast chromosome II carries the carboxypeptidase Y sorting gene PEP1, as well as ACH1, FUS3 and a putative ARS.

AU Van Dyck L; **Purnelle B**; Skala J; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Belgium..  
 SO YEAST, (1992 Sep) 8 (9) 769-76.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 EM 199302

L3 ANSWER 16 OF 20 MEDLINE  
 AN 92327849 MEDLINE  
 DN 92327849  
 TI The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading frames including the RVS161, ADP1 and PGK genes.

AU Skala J; **Purnelle B**; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium..  
 SO YEAST, (1992 May) 8 (5) 409-17.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 EM 199210

L3 ANSWER 17 OF 20 MEDLINE  
 AN 92160395 MEDLINE  
 DN 92160395  
 TI The product of the YCR105 gene located on the chromosome III from Saccharomyces cerevisiae presents homologies to ATP-dependent permeases.

AU **Purnelle B**; Skala J; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium..  
 SO YEAST, (1991 Nov) 7 (8) 867-72.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 EM 199205

L3 ANSWER 18 OF 20 MEDLINE

AN 92116650 MEDLINE  
 DN 92116650  
 TI The open reading frame YCR101 located on chromosome III from  
 Saccharomyces cerevisiae is a putative protein kinase.  
 AU Skala J; **Purnelle B**; Crouzet M; Aigle M; **Goffeau A**  
 CS Unite de Biochimie Physiologique, Universite Catholique de Louvain,  
 Belgium..  
 SO YEAST, (1991 Aug-Sep) 7 (6) 651-5.  
 Journal code: YEA. ISSN: 0749-503X.  
 CY ENGLAND: United Kingdom  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 OS GENBANK-S76380; GENBANK-X62451; GENBANK-X62766; GENBANK-X62767;  
 GENBANK-S70506; GENBANK-S70508; GENBANK-S64911; GENBANK-S64912;  
 GENBANK-S64914; GENBANK-S64915  
 EM 199204

L3 ANSWER 19 OF 20 MEDLINE  
 AN 90366583 MEDLINE  
 DN 90366583  
 TI Novel transport ATPases in yeast.  
 AU **Goffeau A**; Ghislain M; Navarre C; **Purnelle B**;  
 Supply P  
 CS Universite de Louvain, Unite de Biochimie Physiologique,  
 Louvain-la-Neuve, Belgium..  
 SO BIOCHIMICA ET BIOPHYSICA ACTA, (1990 Jul 25) 1018 (2-3) 200-2. Ref:  
 33  
 Journal code: AOW. ISSN: 0006-3002.  
 CY Netherlands  
 DT Journal; Article; (JOURNAL ARTICLE)  
 General Review; (REVIEW)  
 (REVIEW, TUTORIAL)  
 LA English  
 FS Priority Journals; Cancer Journals  
 EM 199012

L3 ANSWER 20 OF 20 MEDLINE  
 AN 83053397 MEDLINE  
 DN 83053397  
 TI Reduction of respiratory-chain cytochrome b by lactate in  
 Saccharomyces cerevisiae.  
 AU Briquet M; **Purnelle B**; Beattie D S; **Goffeau A**  
 SO EUROPEAN JOURNAL OF BIOCHEMISTRY, (1982 Oct) 127 (2) 339-42.  
 Journal code: EMZ. ISSN: 0014-2956.  
 CY GERMANY, WEST: Germany, Federal Republic of  
 DT Journal; Article; (JOURNAL ARTICLE)  
 LA English  
 FS Priority Journals  
 EM 198303

\*\*\*\*\*  
MPSRELA  
(TM)  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPerch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Fri Nov 7 14:58:00 1997; Maspar time 435.94 Seconds  
927.866 Million cell updates/sec  
Tabular output not generated.  
Title: >US-08-822-264-2  
Description: (1-788) from US08822264.seq  
Perfect Score: 784  
N.A. Sequence: 1 GCGCGCCGAACCCGCGCGCC.....AATATTAGAAAGTTTGAGC 788  
Comp: CCGCGCGCTTGGCGCGCGCG.....TTATAATCTTCAACTCG

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:  
EST-STS  
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

Database:  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

Statistics: Mean 10.797; Variance 2.016; scale 5.355

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	449	57.3	520	121	W56474	0.00e+00
2	433	55.2	501	179	AA081900	0.00e+00
3	347	44.3	390	194	AA166645	0.00e+00
4	331	42.2	411	198	H95223	0.00e+00
5	329	42.0	369	27	R59281	0.00e+00
6	328	41.8	438	63	H23045	0.00e+00
7	302	38.5	411	102	N66942	0.00e+00
8	294	37.5	392	120	W49363	0.00e+00
9	283	36.1	401	128	W97563	0.00e+00
10	279	35.6	381	181	AA088127	0.00e+00
11	278	35.5	334	83	H48290	0.00e+00
12	274	34.9	367	155	AA018712	0.00e+00
13	272	34.7	398	120	W50184	0.00e+00
14	269	34.3	330	179	AA080939	0.00e+00
15	268	34.2	359	149	W12485	0.00e+00
16	258	32.9	314	116	W25549	0.00e+00
17	249	31.8	284	86	H60595	0.00e+00
18	241	30.7	406	38	RI3334	0.00e+00
19	226	28.8	274	9	T86596	0.00e+00
20	214	27.3	259	53	R93625	0.00e+00
21	212	27.0	470	9	T86595	0.00e+00
22	211	26.9	296	83	H35803	0.00e+00
23	208	26.5	216	68	H74052	0.00e+00
24	198	25.3	238	38	RI3666	0.00e+00
25	195	24.9	432	68	RI3490	0.00e+00
26	189	23.1	190	14	T09102	0.00e+00
27	179	22.8	260	162	AA103591	0.00e+00
28	167	21.3	177	76	R95452	0.00e+00
29	153	19.5	198	139	AA051109	0.00e+00
30	149	19.0	202	193	AA163568	0.00e+00
31	145	18.5	162	116	W24659	0.00e+00
32	143	18.2	151	38	RI3636	0.00e+00
33	142	18.1	633	188	AA147865	0.00e+00
34	130	16.6	177	131	MMEST725	0.00e+00
35	126	16.1	578	191	AA156043	0.00e+00
36	122	15.6	131	78	R98305	0.00e+00
37	116	14.8	126	27	R56508	0.00e+00
38	112	14.3	288	129	HSC0BE071	0.00e+00
39	100	12.8	449	149	W12817	0.00e+00
40	86	11.0	281	81	T34996	0.00e+00
41	84	10.7	187	81	T36258	0.00e+00
42	82	10.5	358	17	T49867	0.00e+00
43	80	10.4	139	194	AA168924	0.00e+00
44	74	9.2	533	62	H17346	0.00e+00
45	61	7.8	432	162	AA104783	0.00e+00

ALIGNMENTS

RESULT 1 W56474 520 bp mRNA EST 03-JUN-1996  
LOCUS zc59b01.r1 Soares parathyroid tumor Nbhpa Homo sapiens cDNA clone  
DEFINITION 326569 5'  
ACCESSION W56474  
NID g1358332  
KEYWORDS EST.



SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 520)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT  
  
Contact: Wilson RK  
WashU-Merck EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: mob.REGA+ET  
High quality sequence stop: 451.  
Location/Qualifiers  
    1..520  
    /organism="Homo sapiens"  
    /note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5/-]  
TGTTACCAATCTGAAGTGGAGCGGCGGCCACCACAAATTTTTTTTTTTTTTTTTTTTTT  
T-3'}, double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."  
/clone="326569"  
/clone\_lib="Soares parathyroid tumor NbHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
<!--520  
                100 a   160 g   155 g   97 t     8 others  
SE COUNT  
IGIN  
  
Query Match       57.3%; Score 449; DB 121; Length 520;  
Best Local Similarity 94.5%; Pred. No. 0.00e+00;  
Matches 495; Conservative 0; Mismatches 24; Indels 5; Gaps 5;  
  
Db 1 gagaaagtggcgaattccgagtcactgcctgacnccgcccaaccttactccagagatca 60  
QY 41 GAGAAAGTGGCGAGTTCGGGATCCCTGCCTAGCGCGGCCAACCTTTACTCCAGAGATCA 100  
Db 61 tggtcgcgaggatgtggttgcgactggcgc-gaccacaagcgactgtggagagcgcgggc 119  
QY 101 TGGCTGCCGAGGATGTGGTGGCGAGTGGCGCGCAGCCCAAGCGATCTGGAGAGCGCGGGC 160  
Db 120 tgctcgatgatatttcacgtccgcctccaacctgcctgttcgttcgcttcgactctcc 179  
QY 161 TGCTGCATGAGATTTTTCACGTGGCGGTCTCAACCTGCTGCTGTGGCTCTGCGATCTCC 220  
Db 180 tgctctacaagatcgtgcgcgggggaccagccggcgnc-agcgcgacgangacgacgacga 238  
QY 221 TGCTCTACAAGATCGTGGCGGGGACCAGCGCGGGCCAGCGGGCGGACGACGACGACGA 280  
Db 239 -gcncccccctctgcccgccctcaagcg-ogcancnttcccccccgcgagctgcggcnt 296  
|||||

|                       |  |   |  |                   |
|-----------------------|--|---|--|-------------------|
| Qy                    | 281  | NGCCGCCCTCTTGC  | CCCGCTCAAGCGGGCGGACTTCACCCCGCGGAGCTGCGGGCGCT | 340               |
| Db                    | 297  | tcgacggcgctccagagaccgcgcatactactatcgccatcaacggaaggtgttcgatgtga  | 356  |                   |
| Qy                    | 341  | TCGACGGCGTCCAGGACCCGCGCATCTCATATGGCCATCAACGCGAAGGTGTTCGATGTGA   | 400  |                   |
| Db                    | 357  | ccaaagcgcgcaaatctacagggcccgagggcggttanttnggggtctttgctggaagagatg | 416  |                   |
| Qy                    | 401  | CCAAAGGCGCGAAATCTTACGGGGCCGAGGGCGGTATGGGGTCTTTTCTCTGGAAGAGATG   | 460  |                   |
| Db                    | 417  | catcagggcgcttgcacatttgcctgggataaggaagcactgaagatgagtcagatg       | 476  |                   |
| Qy                    | 461  | CATCAGGGGCGTTCGCACATTTTCCTGGATAGGAAGCACTGAAGATGAGTACGATG        | 520  |                   |
| Db                    | 477  | acctttgcactcactgtgctgccagaggagactctgagtga                       | 520  |                   |
| Qy                    | 521  | ACCTTTGACCTCACT-GCTGCCAGCAGGAGACTCTGAGTGA                       | 563  |                   |
| RESULT                | 2  |   |  |                   |
| LOCUS                 | AA081900   | 501 bp  | mRNA   | EST               |
| DEFINITION            | zn23904.r1   | Stratagene neuroepithelium NT2RAMI                              | 937234                                       | Homo sapiens      |
|                       | cDNA clone 548310.5,   | similar to TR:E247050   | E247050                                      | CHROMOSOME XVI    |
|                       | READING FRAME  | ORF YPL170W.  |  |                   |
| ACCESSION             | AA081900   |   |  |                   |
| NID                   | g1623958   |   |  |                   |
| KEYWORDS              | EST.   |   |  |                   |
| SOURCE                | human.   |   |  |                   |
| ORGANISM              | Homo sapiens   |   |  |                   |
| REFERENCE             | Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 501)  |   |  |                   |
| AUTHORS               | Hallier,L., Clark,M., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,K.  |   |  |                   |
| TITLE                 | WashU-Merck EST Project  |   |  |                   |
| JOURNAL               | Unpublished (1995)   |   |  |                   |
| COMMENT               | Contact: Wilson RK<br>WashU-Merck EST Project<br>Washington University School of Medicine<br>4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108<br>Tel: 314 286 1800<br>Fax: 314 286 1810<br>Email: est@wustl.wustl.edu<br>This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.<br>Seq primer: -28M13 rev2 from Amersham<br>High quality sequence stop: 398.<br>Location/Qualifiers<br>1..501<br>/organism="Homo sapiens"<br>/note="vector: phiuscript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2 (Ntera-2/cl.D1) precursor cells induced with Retinoic Acid for 1 week, followed by 3 weeks in mitotic inhibitors (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' /clone="548310"<br>/clone_lib="Stratagene neuroepithelium NT2RAMI 937234"<br>/dev_stage="Ntera-2/RAMi neuroepithelial cells"<br>/lab_host="SOLR (kanamycin resistant)"<br><1..>501 |   |  |                   |
| BASE COUNT            | 96 a   | 155 g   | 92 t   | 7 others          |
| ORIGIN                |  |   |  |                   |
| Query Match           | 55.2%;   | Score 433;  | DB 179;                                      | Length 501;       |
| Best Local Similarity | 95.5%;   | Pred.No. 0.00e+00;  |  |                   |
| Matches               | 484;   | Conservative 0;   | Mismatches 16;                               | Indels 7; Gaps 6; |
| Db                    | 1  | tggcgagttccgagtcctgctgtagcgcgccaccttactacagatcatggctgc          | 60   |                   |



```

Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1536 Std Error: 0.00
Seq primer: M13RP1.
Location/Qualifiers
1..411
/organism="Homo sapiens"
/notes="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGCGGCGCGATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fátima Bonaldo."
/clone="256535"
/clone_lib="Soares placenta 8to9weeks 2NbD8to9w"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
<1..>411
106 a 89 c 113 g 95 t 8 others
mRNA
BASE COUNT

```

| Query Match           | 42.28;          | Score 331;   | DB 198;   | Length 411; |
|-----------------------|-----------------|--|-----------|-------------|
| Best Local Similarity | 95.08;          | Pred. No. 0.00e+00;  |           |             |
| Matches 382;          | Conservative 0; | Mismatches 12;   | Indels 8; | Gaps 3;     |
| Db                    | 14              | cggcgncagagaccgcgcatactcatggccatcnnccggcgaagtggttcgagtgtgaccaa   | 73        |             |
| QY                    | 345             | CGGGTCCAGAGCCCGGCNATACTCATGCCATCAACGGCAAGTGTCGATGTGACCAA         | 404       |             |
| Db                    | 74              | agggccgaattctacggggcccgagggtan - tn - ggggtttcttcgctgggaagatgcac | 131       |             |
| QY                    | 405             | AGGCCGCAAAATTACGGGGCCCGAGGGCGGTATGGGGTCTTTGCTGGGAAGATGCATC       | 464       |             |
| Db                    | 132             | caggggacctgc - acatttgcctgataaggaagcactgaagatgagtagcatgaacct     | 190       |             |
| QY                    | 465             | CAGGGCCCTTGCCACATTTTTCCTGNTAGGAAGCACCTGAAGGATGATGACGATGACCT      | 524       |             |
| Db                    | 191             | ttctgacctcaactgctgcgccagcng - agactctgagtgcactggagctcaggttcac    | 249       |             |
| QY                    | 525             | TTCTGACCTCACCTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTCACTTT     | 584       |             |
| Db                    | 250             | caagtatcatcagctgggcacaaactgctgaaggagggggagagcccaactgtgtactcaga   | 309       |             |
| QY                    | 585             | CAAGTATCATCAGCTGGGCAAACTGCTGAAGAGGGGGAGGAGCCCACTGTGTACTACGA      | 644       |             |
| Db                    | 310             | tgagggaagaccaaagatgagagtgcccggaataatgatataagcatctcagtggaagt      | 369       |             |
| QY                    | 645             | TGAGGAAGAACCAAAAGATGAGAGTTCCCGGAAAAATG - TTAAGCATTT - CAGTGGAGCT | 702       |             |
| Db                    | 370             | atactatttttgtatttttggnaatcatttgtgaagtc                           | 411       |             |
| QY                    | 703             | ATATCTATNNNT - GTATTTT - GCAAAATCATTTGTAAACAGTCC                 | 742       |             |

|            |   |
|------------|---|
| RESULT     | 5   |
| LOCUS      | R59281.1 369 bp mRNA EST 24-MAY-1995  |
| DEFINITION | Y97G07.r1 Homo sapiens cDNA clone 41698 5'.   |
| ACCESSION  | R59281  |
| NID        | 9829976   |
| KEYWORDS   | EST.  |
| SOURCE     | human clone-41698 library=Soares infant brain NTB vector=Lafmid BA<br>host=DH10B (ampicillin resistant) primer=M13Rpl RsiteL=Not I<br>Rsite2=Hind III Whole brain from a 73 days post natal female. 1st<br>strand cDNA was primed with a Not I - oligo(dT) primer [5',<br>AACTGAAGAATTCGGCCGCAGGAATTTTTTTTTTTT 3']; double-stranded<br>cDNA was ligated to Hind III adaptors (Pharmacia), digested with<br>Not I and directionally cloned into the Not I and Hind III sites of<br>the Lafmid BA vector. Library went through one round of |

normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 369)

AUTHORS Hillier,L., Clark,N., Dubucque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson.R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT GDB: G00-414-239  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 295  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

[illegible]

|    |     |   |     |
|----|-----|---|-----|
| Db | 1   | tccttgctagcgcggcccaaccttactccagagatcatgctgcgagagatgctggtg   | 60  |
| QY | 62  | TCCTGCTAGCGCGGCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGGTGG     | 121 |
| Db | 61  | cgaatgagcgc-gacccaacgatctggacagcgcggtctgctgatgatttcaagt     | 119 |
| QY | 122 | CGAATGGCGCGGACCCAGCGATCTGGAGAGCGCGGCTGTGCATGAGATTTTCACT     | 181 |
| Db | 120 | cgcgcgtcaacctgctgctgtggcctctgcatcttctgctctacaagatcgctgcg    | 179 |
| QY | 182 | CGCGCTCAACCTGCTGCTGTGGCTCTGCATCTTCTGCTCTACAGATCGTGGCG       | 241 |
| Db | 180 | ggagaccgcggcggccagcngnagcangacacgacga-gcgcgccctctgcgccgc    | 238 |
| QY | 242 | GGACACAGCGCGCGCGACGGCGACGACACGACGANGCGCGCCCTCTGCCCGCG       | 301 |
| Db | 239 | tcaagcgcgcganccttcaaccgcgcgagctcgcgcttcgacggcgtccaggaaccgc  | 298 |
| QY | 302 | TCAAGCGCGCGGACTTTCACCCCGCGCGCTGCGGCGCTTCGACGCGCTCCAGGACCGC  | 361 |
| Db | 299 | gcatactcatggccatcaacgcgcaaggtttcgatgtnaccaagcgcncaaattctgac | 358 |
| QY | 362 | GCATACTCATGGCCATCAGCGCAAGTGTTCGATGTGACCAAGGCCGCAATCT-AC     | 420 |
| Db | 359 | gggcccgaagg   | 369 |
| QY | 421 | GGGCCCGAGG  | 431 |

| RESULT     | 6 | H23045     | 438 bp       | mrna | EST            | 06-JUL-1995 |
|------------|---|------------|--------------|------|----------------|-------------|
| LOCUS      |   |            |              |      |                |             |
| DEFINITION |   | YMS1G08.11 | homo sapiens | cdna | clone 52059 5' |             |
| ACCESSION  |   | H23045     |              |      |                |             |
| NID        |   | q891740    |              |      |                |             |

EST.  
human clone-52059 library=Soares infant brain lNIB vector=Lafmid BA  
host=DH10B (ampicillin resistant), primer=M13Rp1 Rstel-Not I  
Rst2-Hind III Whole brain from a 73 days post natal female. 1st  
strand cDNA was primed with a Not I - oligo(gt) primer [5',  
AACTGGAAGAAATTCGGCGCCGAGCAATTTTTTTTTTTTTTTT 3']; double-stranded  
cDNA was ligated to Hind III adaptors (Pharmacia), digested with  
Not I and directionally cloned into the Not I and Hind III sites of  
the Lafmid BA vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.Fatima  
Bonaldi.

ORGANISM Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Choroata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 438)  
Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaboa,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
The WashU-Merck EST Project  
Unpublished (1995)

REFERENCE AUTHORS  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

FEATURES  
source  
GDB: G00-424-995  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
High quality sequence stops: 345  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/clone="52059"

BASE COUNT 74 a 137 c 123 g 84 t 20 others  
ORIGIN

Query Match 41.8%; Score 328; DB 63; Length 438;  
Best Local Similarity 92.2%; Pred. No. 0.00e+00;  
Matches 392; Conservative 0; Mismatches 24; Indels 9; Gaps 9;  
b 1 agttccgattccctcctagtcgagcgcccaacctttactccagagatcatggctgcgcgag 60  
QY 53 AGTTCCGATGCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAG 112  
|||||  
Db 61 atgtngtnggactnnggc - gaccacagcagctctngagancgcgggctgctcatgaga 119  
|||||  
QY 113 ATGTGTGTGGCGACATGGCGGCCGACCAACGCGATCTGGAGAGCGCGGCGCTGCTCATGAGA 172  
|||||  
Db 120 ttctcagtcgcccctcaacctgctgctgttggcctctgcatcttctctgctctacaaga 179  
|||||  
QY 173 TTTTCAGCTGCCGCTCAACCTGCTGCTGTGGCCCTTCGATCTTCTGCTCTACAGA 232  
|||||  
Db 180 tctgtcgcggggaccacgcggcgccgagcngancaggacgacgacga - gccgccccctc 238  
|||||  
QY 233 TCGTGC GCGGGACACAGCGCGGCCGACAGCGCGACAGGACGACGACGCGCCGCCCTC 292  
|||||  
Db 239 tnccccgcctaaagctcgcgagantctacccccnccagctcgggntcttncagcggcgt 298  
|||||  
QY 293 TGC CCGCCCTCAACGCGCGGA - CTTACCCCCCGGAGCTGCGGCGCTT - CGACGCGCT 350  
|||||  
Db 299 ccaggnaccgncgcatcactcatggcccatcacgcgcaagtgttcgatgacgacaaaggc 358  
|||||  
QY 351 CCAGG - ACCCGCGCATACTCATNG - CCATCAACGGCAAGGTTCGATGTGACCAAGGC 408  
|||||  
Db 359 cgnaaattctacgngnccgagggcgtttattgggntactttgctnggaangatgcattcc 418

| QY                    | 409   | CGCAAAATCTACGGCCCGAGGGCCGATGGGG-T-CTTTGCT-GGAAGAGATGCATCC | 465         |
|-----------------------|---|---|-------------|
| Db                    | 419   | agggg 423   |             |
| QY                    | 466   | AGGG 470  |             |
| RESULT                | 7   |   |             |
| LOCUS                 | N66942  | 411 bp mRNA   | EST         |
| DEFINITION            | za48cl2.s1 Homo sapiens cDNA clone 295798 3'  |   | 08-MAR-1996 |
| ACCESSION             | N66942  |   |             |
| NID                   | g1219067  |   |             |
| KEYWORDS              | EST.  |   |             |
| SOURCE                | human   |   |             |
|                       | clone=295798 primer=ml3 -40 forward library=Soares fetal liver spleen INPLS vector=p7T73D (Pharmacia) with a modified polylinker host=DHI0B (ampicillin resistant) Reitel-Pac I Reite2-Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(df) primer [5' - AACTGGAGAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p7T73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. |   |             |
| ORGANISM              | Homo sapiens  |   |             |
|                       | Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcosterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoide; Homo. 1 (bases 1 to 411)   |   |             |
| REFERENCE             | 1 (bases 1 to 411)  |   |             |
| AUTHORS               | Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,E., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.   |   |             |
| TITLE                 | The WashU-Merck EST Project   |   |             |
| COMMENT               | Unpublished (1995)  |   |             |
|                       | Contact: Wilson RK  |   |             |
|                       | WashU-Merck EST Project   |   |             |
|                       | Washington University School of Medicine  |   |             |
|                       | 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108   |   |             |
|                       | Tel: 314 286 1800   |   |             |
|                       | Fax: 314 286 1810   |   |             |
|                       | Email: est@watson.wustl.edu   |   |             |
|                       | High quality sequence stops: 365  |   |             |
|                       | Source: IMAGE Consortium, LLNL  |   |             |
|                       | This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.   |   |             |
| FEATURES              | Location/Qualifiers   |   |             |
| source                | 1..411  |   |             |
|                       | /organism="Homo sapiens"  |   |             |
| mRNA                  | /clone="295798"   |   |             |
| BASE COUNT            | 114 a 106 c 74 g 116 t 1 others   |   |             |
| ORIGIN                | <1..>411  |   |             |
| Query Match           | 38.5%; Score 302; DB 102; Length 411;   |   |             |
| Best Local Similarity | 97.9%; Pred. No. 0.00e+00;  |   |             |
| Matches               | 321; Conservative 0; Mismatches 4; Indels 3; Gaps 3;  |   |             |
| Db                    | 1 aagcacagtgagctgttacaaatgattttgcacaaatagatactatccact 60  |   |             |
|                       |   |   |             |
| Cp                    | 752 AAGCAGANGTGAGCTGTACAAATGATTTGCAAAATACANNA-TAGATATATCTCCACT 694  |   |             |
|                       |   |   |             |
| Db                    | 61 gaatgctttaatcattttccggggcactctcatcttttggcttctctcatctgagta 120  |   |             |
|                       |   |   |             |
| Cp                    | 693 GAATGCTTAA-CATTTTCCGGGAACCTCATCTTTGGTCTTCCCTCATCTGAGTACA 635  |   |             |
|                       |   |   |             |
| Db                    | 121 cagtgggtctctccccctcttcagcagtgatggccacgtgatgactgaagtgaact 180  |   |             |
|                       |   |   |             |
| Cp                    | 634 CAGTGGGCTCTCCCTCTCTTCAGCACTTGCACAGTTCGCCACGTGATGACTTGAAGTGA 575   |   |             |
|                       |   |   |             |

```

Db 181 gagactccagtcactcagagtcctctgtggtgagcagcagtcaggtcagaagaagtcacgtc 240
|||||
Cp 574 GAGACTCCAGTCACACTCAGAGTCTCTGCTGGCGCAGCAGTGAAGTCAAGATCATCGT 515
|||||
Db 241 actcctcctcagtccttctctatccagcaaaatgtgcaagccc-tgagtcacatctc 299
|||||
Cp 514 ACTCATCTCTCAGTGTCTCTTATCATCAGGCAAAATGTGGCAAGCCCTCGATGCAATCTC 455
|||||
Db 300 ttccagcaagaagcccccatacagccctc 327
|||||
Cp 454 TTCCAGCAAGACCCCATACGGCCCTC 427
|||||

RESULT 8
LOCUS W49363 392 bp mRNA EST 28-MAY-1996
DEFINITION md29b03.r1 Life Tech mouse brain Mus musculus cDNA clone 369773 5'.
ACCESSION W49363
NID 91337638
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 392)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:231205
Seq primer: EnPrimer
High quality sequence stop: 326.
Location/Qualifiers
1..392
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pCMV-SPORT2; site_1: SalI; site_2: NotI;
Cloned unidirectionally. Primer: Oligo dt. pCMV-SPORT2
vector."
/clone="369773"
/clone_lib="Life Tech mouse brain"
/dev_stage="adult"
/lab_host="DH10B"
<1..>392
MRNA 73 a 128 c 120 g 71 t
BASE COUNT
ORIGIN
Query Match 37.5%; Score 294; DB 120; Length 392;
Best Local Similarity 88.8%; Pred.No. 0.00e+00;
Matches 348; Conservative 0; Mismatches 42; Indels 2; Gaps 2;

Db 2 ggaggaagccgactgttcggatctctgcatagcagggcccaaccttctgcacagatc 61
|||||
Qy 40 GGAGAAAGTGGCGAGTTCGGATCCCTGCTAGCGGGCCCAACCTTACTCCAGAGATC 99
|||||
Db 62 atggtgcagagatgtgtgagcactggtgcgcagccccagcagcagtcagggtgcggcg 121
|||||
Qy 100 ATGGCTGGCGAGATGTGGTGGGACTGGCGCGGACCCAGCCAGCATCTGGAGAGCGCGGG 159
|||||

```

```

Db 122 ctcttcaagagattttcaagtcgctctcaactgctgctccttggcctctgcatcttc 181
|||||
Qy 160 CTGCTGATGAGATTTTCAGTCGCGCTCAACCTGCTGCTGTGTCCTTGCATCTTC 219
|||||
Db 182 ctgctctacaagatcgttcgctggggaccagcccggtgctcagtggtgggagacagcagac 241
|||||
Qy 220 CTGCTCTACAGATCTGTCGCGGGGACCAGCGCGGCGGCGGCGGACAG-GAGGAGGAC 278
|||||
Db 242 ga-gcgccccctgctgccccctcaagcgctgacttcacccccctcccaataagcgc 300
|||||
Qy 279 GANCGCGCCCTCTGCGCGCTCAAGCGCGGCGACTTCACGCCCGCGAGGTGCGGCG 338
|||||
Db 301 atcagtgagtcacagaccgctcttctatggccatcaacggcaaggtgttcacgt 360
|||||
Qy 339 CTTTCGAGGGGCTCCAGGACCGCGCATCTCATGGCCATCAACGGCAAGGTTCGATGT 398
|||||
Db 361 gaccaagcgccgaagtctctatggcgcgag 392
|||||
Qy 399 GACCAAGGCGCAATTCACGGGCGCGAGG 430
|||||

RESULT 9
LOCUS W97563 401 bp mRNA EST 16-JUL-1996
DEFINITION mg02a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus CDNA
clone 422570 5'.
ACCESSION W97563
NID 91427684
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 401)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:257122
Seq primer: mob.REGA+ET
High quality sequence stop: 358.
Location/Qualifiers
1..401
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
/clone="422570"
/sex="unknown"

```

```

/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
<1..>401
BASE COUNT      70 a 136 c 124 g 71 t
ORIGIN

Query Match      36.1%; Score 283; DB 128; Length 401;
Best Local Similarity 87.9%; Pred. No. 0.00e+00;
Matches 350; Conservative 0; Mismatches 44; Indels 4; Gaps 4;

Db 7 ggcggaggagagactgttccaggactgtcctagcagcggcccaaccttggctccagag 66
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 37 GGAGGAGAAAGTGGCGAGTTCGGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAG 96
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 67 atcatgtgcccagagatggtgggactggcg-c-gaccgcgagcgtcagaggcg-c 124
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 97 ATCATGGCTGCCGAGATGTTGGGACTTGGCGCCGACCAAGCGATCTGGAGAGCGGC 156
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 125 gggctgtcacagagatgtttcacgtctctctcaacctgtctctctctggcctctgcatc 184
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 157 GGGCTGCTGCATGAGATTTTCACGTGCGCGCTCAACCTGCTGCTGTGGCTTGCATC 216
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 185 ttctctctacaagatgcttcgaggaccagcccggtgcagtgatcaaacagcagac 244
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 217 TTCTCTCTCTACAGATCTGTCGCGGGGACAGCGCGCGCGCGGACAG-GAGGAC 275
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 245 gacgaa-cacccccctgcccgcctcaagcggcgatcttcaacctccagcagctgag 303
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 276 GAGGANGCGCCGCCCTCTGCGCCGCTCAAGCGCGCGACTTACCCGCCGCGAGCTGCG 335
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 304 gggcttcgatgggtccagagaccgagcattctcatgccatcaacggcgaaggtgttcga 363
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 336 GCGCTTCGACGGGCTCCAGGACCGCGCATCTCATGCGCATCAACGGCAAGGTGTCGA 395
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 364 cgtgaccaagcgccgaagtctacggcctgaggggc 401
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 396 TGTGACAAAGCGCGAAATTTCTACGGCGCGGAGGGC 433
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

RESULT 10
LOCUS      AA088127      381 bp      mRNA      23-OCT-1996
DEFINITION      m89a10.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
551226 5', similar to TR:E247050 E247050 CHROMOSOME XVI READING
FRAME ORF YPL170W. ;.
ACCESSION      AA088127
NID            91629719
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
               Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Mus.
REFERENCE      1 (bases 1 to 381)
AUTHORS        Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
               Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
               Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
               Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
               Waterston, R.
TITLE          The WashU-HMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        Contact: Marra M/Mouse EST Project
               WashU-HMI Mouse EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: mouseest@watson.wustl.edu
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               MGI:332018
               Possible reversed clone: similarity on wrong strand

Seq primer: primer name ambiguous
High quality sequence stop: 336.
Location/Qualifiers
1..381
/organism="Mus musculus"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. M30 CD4+
cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
/clone="551226"
/clone_lib="Stratagene mouse Tcell 937311"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
<1..>381
BASE COUNT      65 a 131 c 115 g 70 t
ORIGIN

Query Match      35.6%; Score 279; DB 181; Length 381;
Best Local Similarity 89.2%; Pred. No. 0.00e+00;
Matches 339; Conservative 0; Mismatches 37; Indels 4; Gaps 4;

Db 5 gttccggagctctcctagcggcccaaccttggctccagagatcgtgctccagagga 64
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 54 GTTCCGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGTGGCGAGGA 113
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 65 tgtgtgagcactggcg-c-gaccgcgagcgtcagaggcgcggtgctgcacagat 123
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 114 TGTGTTGGGACTGGCGCGACCAAGCGATCTGGAGAGCGGGGCTGCTGATGAGAT 173
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 124 ttcaactgtctctcaacctgtctctctggcctctgcatcttctctgtctcaaat 183
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 174 TTTTACGTCGCGCTCAACCTGCTGCTGTTGGCTCTGCTGCTCTTCTGCTCTACAAGAT 233
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 184 cgttcggggaccagcccggtgcagtgacgaacacgacgacgacgaa-caccccctcct 242
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 234 CGTGGCGGGGACGACCGCGCGCGAGGAGGACGACGANGCGCGCCCTCT 293
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 243 gccccctcaagcg-cgcgacttccacctcgcagctcgcagctgagcgttcgagcgtcc 301
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 294 GCCCGGCTCAAGCGCGCGACTT-CACCCCGCGAGCTCGCGGCTTCAGCGCGTCC 352
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 302 aggaccgcgctctcattcgtccatcaacggcgaaggttctcagctgaccaaagcccgca 361
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 353 AGGACCGCGATACTCATGCCATCAACGGCAAGGTTTCGATGTGACCAAGGCCGCA 412
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 362 agttctacggcctgagggg 381
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 413 AATTCTACGGGCGCGAGGGG 432
  ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

RESULT 11
LOCUS      H48290      334 bp      mRNA      EST      14-SEP-1995
DEFINITION      Y669407.s1 Homo sapiens cDNA clone 201084 3'.
ACCESSION      H48290
NID            9986677
KEYWORDS      EST.
SOURCE        human clone-201084 primer-Promega -21ml3 library-Soares fetal liver
               spleen INFLS vector-pt7T3b (Pharmacia) with a modified polylinker
               host-DH10B (ampicillin resistant) Reitel-Pac I Rsite2-Eco RI Liver
               and spleen from a 20 week-post conception male fetus. 1st strand
               cDNA was primed with a Pac I - oligo(dt) primer [5',
               AACTGAGAGATTAATTAAGATCTTTTATTTTATTTT 3'], double-stranded
               cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
               I and cloned into the Pac I and Eco RI sites of the modified p7T3
               vector. Library went through one round of normalization. Library
               constructed by Bento Soares and M. Fatima Bonaldi.
ORGANISM      Homo sapiens
               Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
               Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
               Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
               Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 334)

```

**AUTHORS** Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

**TITLE** The WashU-Merck EST Project

**JOURNAL** Unpublished (1995)

**COMMENT** Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 289

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

**FEATURES** Location/Qualifiers

source

1..334

/organism="Homo sapiens"

/clone="201084"

<1..>334

BASE COUNT 82 a 91 c 70 g 88 t 3 others

ORIGIN

Query Match 35.5%; Score 278; DB 83; Length 334;

Best Local Similarity 95.8%; Pred. No. 0.00e+00;

Matches 316; Conservative 0; Mismatches 8; Indels 6; Gaps 6;

Db 1 aagacagagtgacgttacaatgatatttncacaaatacaaaatagatacttccact 60

Cp 752 AAGACANAGTGACGTTCACAAATGATTTGCAAAATACANNA-TAGATATATCTCCACT 694

Db 61 gaatgctttaatacattttncggggaactctcatcttttgggtttcttctcatctgagtaga 120

Cp 693 GAATGCTTTAA-CAATTTTCCGGGAACCTCATCTCTTTTGGTTCTTCTCATCTAGTAC 635

Db 121 cagtggtgctctccctctcctcagcagtttccacagtgatgatacttgaaagtgaact 180

Cp 634 CAGTGGGCTCTCTCCCTCTCTCAGCAGTTCGCCACGTGATGATCTTGAAGTGAAT 575

Db 181 gagactccagtcactcagagtcctctgctgggacagtcagtgacagaagtcacgt 240

Cp 574 GAGACTCCAGTCATCAGAGTCTCTGCTGGCAGCAGTCAGTGGTCAAGGCTATCGT 515

Db 241 actcatcttcagtgcttctctatccaggggcaaaatgtgggcaagccctggagtcacat 300

514 ACTCATCTCTCAGTGCTTCTTATCCAGG-CAAAATGTGG-CAAGGCCCTGG-ATGCAT 458

Db 301 ctcttccagcaaaagccccacacagggccc 330

Cp 457 CTCCTCCAGCAAG-ACCCCATACGGCCCC 429

**RESULT** 12

LOCUS AA016712 367 bp mRNA EST 02-AUG-1996

DEFINITION mg90h03.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA

clone 440309 5'.

ACCESSION AA016712

NID g1478943

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae;

Murinae; Mus

1 (bases 1 to 367)

**REFERENCE**

**AUTHORS** Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

**TITLE** JOURNAL

**COMMENT**

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:265645

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 275.

**FEATURES** source

Location/Qualifiers

1..367

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT

T 3'/], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru KO, Wayne

State Univ., from 2 ]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo."

/clone="440309"

/clone\_lib="Soares mouse embryo NBME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

<1..>367

mRNA

BASE COUNT 63 a 127 c 111 g 66 t

ORIGIN

Query Match 34.9%; Score 274; DB 155; Length 367;

Best Local Similarity 89.8%; Pred. No. 0.00e+00;

Matches 326; Conservative 0; Mismatches 35; Indels 3; Gaps 3;

Db 7 ctgagcggggcccaacttctccagagatcatggtgcccagagatgtgtgagcactg 66

QY 69 CTAGCGGGGCCCAACCTTTACTCCAGAGATCATGGCTGCCGAGGATGTGTGGCGACTGG 128

Db 67 cgc-gacccgagcagtagagggcgggcggtgctgcacgagatttcacgtctctct 125

QY 129 CGCCGACCCCAAGCGATCTGGAGAGCGCGGGCTGCTGCATGAGATTTCACGTGCGCGCT 188

Db 126 caacttgcctctctggcctctgcatcttctctgctctcaagatgcttcgcggggagacca 185

QY 189 CAACCTGCTGCTGCTTGGCTCTGCATCTCTCTGCTCTACAGATCGTGCAGGAGACCA 248

Db 186 gccggtgccaagtcgacacacgacgacgagaa-ccaccccccgctgcccgccctcaagcg 244

QY 249 GCGGGCGGCGCAGGGCGGACGAGGACGACGAGANGCGCCCTCTGCGCCGCGCTCAAGCG 308

Db 245 -cgcgacttcaacctgcccagcagctgagggcttccgctccagcagcccgccgactct 303

QY 309 GCAGCATCTTACCCCGCCGAGCTGCGGGCTTTCAGCGGCGTCCAGGACCCGCGCATACT 368

Db 304 catggccatcaacggcaagtgcttcacgtgacccaagccgcaagttctacgggctga 363

QY 369 CATGGCCATCAACGGCAAGGTGTCATGTGACCAAGGCCCAAAATCTACGGGCCCGA 428

Db 364 ggagg 367

QY 429 GGGG 432

RESULT 13  
 LOCUS W50184 398 bp mRNA EST 29-MAY-1996  
 DEFINITION mb08c05.r1 Life Tech mouse brain Mus musculus cDNA clone 319592 5'.  
 ACCESSION W50184  
 NID g1338602  
 KEYWORDS EST.  
 SOURCE mouse mouse.  
 ORGANISM Mus musculus  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
 Murinae; Mus.  
 REFERENCE 1 (bases 1 to 398)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT

Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:210208  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 318.  
 Location/Qualifiers  
 1..398  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /note="Vector: pCMV-SPORT2; Site\_1: Sali; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. pCMV-SPORT2 vector."  
 /clone="319592"  
 /clone\_lib="Life Tech mouse brain"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 <1..>398 76 a 126 c 122 g 74 t

## FEATURES

source  
 1..398  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /note="Vector: pCMV-SPORT2; Site\_1: Sali; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. pCMV-SPORT2 vector."  
 /clone="319592"  
 /clone\_lib="Life Tech mouse brain"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 <1..>398 76 a 126 c 122 g 74 t  
 mRNA  
 BASE COUNT 76 a 126 c 122 g 74 t  
 ORIGIN  
 Query Match 34.7%; Score 272; DB 120; Length 398;  
 Best Local Similarity 87.8%; Pred. No. 0.00e+00;  
 Matches 351; Conservative 0; Mismatches 43; Indels 6; Gaps 6;

Db 3 agagggcgaggagcagcagctgttcggatctctgcatagca-ggcccacatttgcctc 61  
 QY 33 AGAGGGAGGAGAAAGTGGCGAGTTCCGGATCCCTGCTAGCGGGCCCCAACCTTACFCC 92  
 Db 62 agagatcatggtccagagagatggtggtgagctggcg-gaccccagcagctggagg 120  
 QY 93 AGAGATCATGCTCCGAGGATGTTGGTGGCAGCTGGCGCCGACCAAGCGATCTGGAG 152  
 Db 121 cg-cgggctctcaagagatttcacgtgcctctcaacctgtgctccttgcccttg 179  
 QY 153 CGCGGGCTGCTGTCATGAGATTTTACGTCGCCGCTCAACCTGCTGCTGTGGCCTCTG 212  
 Db 180 catcttctgctctacaagatcgcttcgggggacagcccggtgacagtgagggaacaga 239  
 QY 213 CATCTTCTCTCTACAGATCGTGCAGGGGACAGCCGCGGCGCCAGCGGACAG-GA 271  
 Db 240 cgacgacga-gccgccccccgtgccccgcctcaagaggtgtcaottcaccctcgccaa 298  
 QY 272 CGAGCAGGANGCCGCCCTCTGCCCGCCCTCAAGCGGC-GCGACTTCAACCCGCCCGC 330

Db 299 ctaaggcgatcacgatggagtcaggaccgcgcattcttattggccatcaacggaaggtg 358  
 QY 331 CTGGCGCGCTTCGAGCGGCTCCAGGACCGCGCATCTATGCGCATCAACGGCAGGTG 390  
 Db 359 ttacacgtgacaaagccgcaagtcttatggcgagg 398  
 QY 391 TTCGATGTGACAAAGCGCGCAAAATCTACGGGCCGAGG 430  
 RESULT 14  
 LOCUS AA080939 330 bp mRNA EST 21-OCT-1996  
 DEFINITION zn18a05.r1 Stratagene neuroepithelium NT2RAMI 937234 Homo sapiens  
 cDNA clone 547760 5' similar to TR:E247050 E247050 CHROMOSOME XVI  
 READING FRAME ORF YPL170W. ;  
 ACCESSION AA080939  
 NID g1623638  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 330)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasaki, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 TITLE WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28M13 rev2 from Amersham.  
 Location/Qualifiers  
 1..330  
 /organism="Homo sapiens"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2 (Ntera-2/cl.D1) precursor cells induced with Retinoic Acid for 1 week, followed by 3 weeks in mitotic inhibitors (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3', adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."  
 /clone="547760"  
 /clone\_lib="Stratagene neuroepithelium NT2RAMI 937234"  
 /dev\_stage="Ntera-2/RA+NI neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 <1..>330 69 a 97 c 94 g 62 t 8 others

## BASE COUNT

ORIGIN  
 Query Match 34.3%; Score 269; DB 179; Length 330;  
 Best Local Similarity 92.4%; Pred. No. 0.00e+00;  
 Matches 306; Conservative 0; Mismatches 21; Indels 4; Gaps 4;  
 Db 1 gacangcngacgacgagcggccctctgcccgcctcaagcg-cgcgagctcacccc 59  
 QY 265 GACAGGACGACGACGANGCGGCCCTCTGCCGCCCTCAAGCGCGGCAC-TTCACCCC 323  
 Db 60 cgcgagctgctgggcnttcgagcggtccaggaccgcgcatactcatggccatcaacgg 119  
 QY 324 CGCGAGGTGGCGGCTTCGAGCGGTTCAGGACCGCGGCATCTCATGGCCATCAACGG 383  
 Db 120 caaggtgttcgatgtgaccaaagccgcaaatcttacggcccgngngtagtn-ggggt 178  
 QY 384 CAAGGTGTTCGATGTGACAAAGCGCGCAAAATTCACGGGCCGAGGGCCGTATGGGT 443





## KEYWORDS

SOURCE human clone-52059 library-Soares infant brain lN18 vector-Lafmid BA host-DH10B (ampicillin resistant) primer-M13R1 Rsite1-Not I Rsite2-Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGATTCGCGCGCAGGAGATTTTITTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcoterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 438)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

## TITLE

The WashU-Merck EST Project

## JOURNAL

Unpublished (1995)

## COMMENT

GDB: G00-424-995  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
High quality sequence stops: 345  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

source 1..438  
/organism="Homo sapiens"  
/clone="52059"  
BASE COUNT 74 a 137 c 123 g 84 t 20 others  
ORIGIN  
Query Match 41.8%; Score 328; DB 63; Length 438;  
Best Local Similarity 92.2%; Pred. NO. 0.00e+00;  
Matches 392; Conservative 0; Mismatches 24; Indels 9; Gaps 9;  
b 1 agttccggatccctgctagtcgagcgcccaacatttactccagagatcatggtcgccgagg 60  
53 AGTTCGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGATCATGCTGCCGAGG 112  
Db 61 atgtngtcgactngcgc-gaccacaagcgtatcngagancgagcggtgctgcatgaga 119  
QY 113 ATGTGTGGGACGTGGCGCGCCGACCAACGCGATCTGGAGAGCGCGGCTGTGATGAGA 172  
Db 120 ttctacgtgcgcgtcaaacctgtgctgttggcctctgcatcttctctctacaga 179  
QY 173 TTTTCAGTCCGCGCTCAACTGCTGTGCTGTGGCTTGTGATCTTCTGCTCTACAGA 232  
Db 180 tcgtgcgaggacacccgagcgccgagcngancagagacgacgacga-gccgccccctc 238  
QY 233 TCGTGGCGGGACACCGCGCGCCAGCGGCGGACAGGACGACGACGANGCGCCCTC 292  
Db 239 tnccegcctcaagcgtcgcganccttccaccccccagcgtcgngntcttncgacggcgt 298  
QY 293 TGCCCCGCCCTCAAGCGCGCGGA-CTTCAACCCCGCGAGCTGCGCGCGCTT-CGACGGCGT 350  
Db 299 ccaggnaccngncatctactatggtggcccatcaacggaaggtgttcgattgtgacaaaggg 358  
QY 351 CCAGG-ACCCGCGCATACTCATGG-CCATCAACGGCAAGGTGTTTCGATGTACCAAGGC 408  
Db 359 cgnaaattctacgggncggagcgctttatttgggntactttgctnggaagngatgcatcc 418

QY 409 CGCAATTTCTAGGCGCCCGAGGGCGGTATGGGG-T-CTTTGCT-GGAAGAGATGCATCC 465  
Db 419 agggg 423  
QY 466 AGGGG 470

## RESULT

LOCUS N66942 411 bp mRNA EST 08-MAR-1996  
DEFINITION za48c12.sl Homo sapiens cDNA clone 295798 3'.  
ACCESSION N66942  
NID 91219067  
KEYWORDS EST.  
SOURCE human clone-295798 primer-m13 -40 forward library-Soares fetal liver spleen lN18 vector-pT7T3D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) Rsite1-Pac I Rsite2-Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGATTAATTAAGATCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcoterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 411)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

## TITLE

The WashU-Merck EST Project

## JOURNAL

Unpublished (1995)

COMMENT  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
High quality sequence stops: 365  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

source 1..411  
/organism="Homo sapiens"  
/clone="295798"  
BASE COUNT 114 a 106 c 74 g 116 t 1 others  
ORIGIN  
Query Match 38.5%; Score 302; DB 102; Length 411;  
Best Local Similarity 97.9%; Pred. NO. 0.00e+00;  
Matches 321; Conservative 0; Mismatches 4; Indels 3; Gaps 3;  
Db 1 aagacagagtggactgttacaaatgatttgcacaaatacaaaatagatatattccact 60  
Cp 752 AAGACANAGTGGACGTGTACAAATGATTTTGCAAAATACANNA-TAGATATATCTCCACT 694  
Db 61 gaatgcttaacatctttccgggacgtctcatctcttggcttctctctctctgagaca 120  
Cp 693 GAATGCTTTAA-CATTTTCCGGGAACCTCATCTTTGGTCTCTCTCATCTGAGTACA 635  
Db 121 cagtggtgctctcccccctctcagcaggttggccacgctgagatgacttgaaagtgaact 180  
Cp 634 CAGTGGGCTCTCCCTCCCTTCAGCAGTTTGGCCACGTGATGATACITGAAAGTGAAC 575



(TM)

| Result No. | Score | Query |   | Length | DB | ID     | Description           | Pred. No. |
|------------|-------|-------|---|--------|----|--------|-----------------------|-----------|
|            |       | Match |   |        |    |        |                       |           |
| 1          | 90    | 5     | 8 | 64     | 10 | R50301 | Drosophila anti-bacte | 2.11e+01  |
| 2          | 90    | 5     | 8 | 64     | 10 | R50302 | Drosophila anti-bacte | 2.11e+01  |
| 3          | 90    | 5     | 8 | 64     | 10 | R50293 | Anti-bacterial glycop | 2.11e+01  |
| 4          | 89    | 5     | 8 | 157    | 17 | R77667 | Corticotropin releasi | 2.50e+01  |
| 5          | 89    | 5     | 8 | 393    | 21 | W13392 | Lipase showing homolo | 2.50e+01  |
| 6          | 89    | 5     | 8 | 393    | 21 | W13391 | Lipase showing homolo | 2.50e+01  |
| 7          | 89    | 5     | 8 | 411    | 21 | W16481 | Human corticotrophin  | 2.50e+01  |
| 8          | 89    | 5     | 8 | 411    | 15 | R50576 | Rat CRF2-alpha recept | 2.50e+01  |
| 9          | 89    | 5     | 8 | 411    | 15 | R50576 | Human CRF2 receptor.  | 2.50e+01  |
| 10         | 89    | 5     | 8 | 420    | 21 | W20266 | H. pylori transmembra | 2.11e+01  |
| 11         | 90    | 5     | 8 | 424    | 21 | W20977 | H. pylori transmembra | 2.11e+01  |
| 12         | 89    | 5     | 8 | 431    | 17 | R57293 | Mouse CRF RBL recepto | 2.50e+01  |
| 13         | 89    | 5     | 8 | 431    | 15 | R50575 | Rat CRF2-beta recepto | 2.50e+01  |
| 14         | 88    | 5     | 7 | 289    | 12 | R66796 | Novel mouse proteogly | 2.96e+01  |
| 15         | 88    | 5     | 7 | 311    | 17 | R37001 | Mouse syndecan-1.     | 2.96e+01  |
| 16         | 88    | 5     | 7 | 311    | 12 | R66793 | Novel mouse proteogly | 2.96e+01  |
| 17         | 88    | 5     | 7 | 311    | 10 | R55276 | Syndecan protein.     | 2.96e+01  |
| 18         | 88    | 5     | 7 | 410    | 12 | R69555 | Human lysosomal membr | 2.96e+01  |
| 19         | 88    | 5     | 7 | 419    | 19 | W01504 | Wild-type human pancr | 2.96e+01  |
| 20         | 88    | 5     | 7 | 419    | 19 | W01509 | Human pancreatic carb | 2.96e+01  |

[illegible]

|     |   |
|-----|---|
| DE  | Lipase showing homology to Lp-PLA2.         |
| KW  | Lipoprotein associated phospholipase        |
| KW  | platelet activating factor; atherosclerosis |
| KW  | endothelial dysfunction; myocardial         |
| KW  | cancer; acute chronic inflammation;         |
| KW  | neuropsychiatric illness; rheumatoid        |
| KW  | human; benign possible hyperplasia          |
| OS  | Homo sapiens.                               |
| Key | Location/Qualifiers                         |
| FF  | Misc_difference 23                          |
| FT  | /note= "encoded by GATG"                    |
| FT  | /note= "encoded by 27                       |
| FT  | Misc_difference 27                          |
| FT  | /note= "encoded by TCAG"                    |
| FT  | /note= "encoded by 71                       |
| FT  | Misc_difference 71                          |
| FT  | /note= "encoded by ACG"                     |
| FT  | /note= "encoded by 72                       |
| FT  | Misc_difference 72                          |
| FT  | /note= "encoded by ACT"                     |
| FT  | /note= "encoded by 97                       |
| FT  | Misc_difference 97                          |
| FT  | /note= "encoded by CACM"                    |
| FT  | /note= "encoded by 99                       |
| FT  | Misc_difference 99                          |
| FT  | /note= "encoded by GGAC"                    |
| FT  | /note= "encoded by 102                      |
| FT  | Misc_difference 102                         |
| FT  | /note= "encoded by TRAC"                    |
| FT  | /note= "encoded by 104                      |
| FT  | Misc_difference 104                         |
| FT  | /note= "encoded by ATTG"                    |
| FT  | /note= "encoded by 106                      |
| FT  | Misc_difference 106                         |
| FT  | /note= "encoded by AGTC"                    |
| FT  | /note= "encoded by 127                      |
| FT  | Misc_difference 127                         |
| FT  | /note= "encoded by TACA"                    |
| FT  | /note= "encoded by 189                      |
| FT  | Misc_difference 189                         |
| FT  | /note= "encoded by CCGG"                    |
| FT  | /note= "encoded by 248                      |
| FT  | Misc_difference 248                         |
| FT  | /note= "encoded by GGCC"                    |
| FT  | /note= "encoded by 251                      |
| FT  | Misc_difference 251                         |
| FT  | /note= "encoded by GACC"                    |
| FT  | /note= "encoded by 254                      |
| FT  | Misc_difference 254                         |
| FT  | /note= "encoded by CTCG"                    |
| FT  | /note= "encoded by 389                      |
| FT  | Misc_difference 389                         |
| FT  | /note= "encoded by CCAT"                    |
| FT  | /note= "encoded by 12984 -Al.               |
| PN  | WO9712984 -Al.                              |
| PD  | 10-APR-1997.                                |
| PF  | 29-SEP-1995; G02320.                        |

CC stroke, diabetes and neuropsychiatric illnesses. The present sequence  
CC and the polypeptide can also be used for antibody or transgenic animal  
CC production.  
CC N.B. The sequence as printed, and as reproduced here, does not  
CC decode directly from the cDNA sequence shown in the specification  
CC (see features table).  
SQ Sequence 393 AA;

Db 331 kfstetrgslpbyegevmvrmlaf-lqkhlldkedynqwnllleglgspl t 383  
||::| | : : : |::| | |::| | : : :  
Best Local Similarity 23.9%; Fred. NO. 2.50E+01;  
Matches 14; Conservative 19; Mismatches 16; Indels 5; Gaps 4

QY 105 KFYGPE--GPGVFAGRDA-SRGLATECLDKB-ALKDEYDDLSDLTAAQQTLS 154

RESULT 7  
ID W16481 standard; Protein; 411 AA.

AC W16481;  
DE 20-JUN-1997 (first entry)  
DE Human corticotropin releasing factor 2 receptor protein.  
KW Human; corticotropin; corticotropin; releasing factor 2; CRF2;  
KW receptor; screen; agonist; antagonist; activation; inhibition;  
KW prevention; treatment; dementia; obesity; acceleration;  
KW stress adaptation; melancholia; anxiety; stress headache; AIDS;  
KW acquired immunodeficiency syndrome; Alzheimer's disease;  
KW gastrointestinal disorder.  
OS Homo sapiens.  
PN J09070289-A.  
PD 18-MAR-1997.  
PF 14-SEP-1995; 237081.  
PR 27-JUN-1995; JP-161213.  
PR (TAKE ) TAKEDA CHEM IND LTD.  
PR WPI; 97-230023/21.  
DR N-PSDB; T66508.  
PT PCR primer for G protein conjugate type receptor protein DNA - and  
PT human corticotropin releasing factor 2 receptor protein, useful to  
PT screen for agonists and antagonists to treat dementia and anxiety  
PS Claim 8; Pages 39-40; 46pp; Japanese.  
CC The present sequence is the human corticotropin releasing  
CC factor 2 (CRF2) receptor protein, which can be used to screen for  
CC an agonist or antagonist which activates the receptor, or  
CC competitively inhibits the binding of the receptor to CRF. The  
CC agonist can be used to prevent or treat dementia and obesity, or  
CC accelerate stress adaptation. The antagonist can be used to prevent  
CC or treat melancholia, anxiety, stress headaches, AIDS, Alzheimer's  
CC disease or gastrointestinal disorders.  
SQ Sequence 411 AA;

Query Match 5.8%; Score 89; DB 21; Length 411;

Best Local Similarity 41.9%; Pred. No. 2.50e+01;

Matches 13; Conservative 9; Mismatches 8; Indels 1; Gaps 1;

Db 259 epgdldvdygqpllvlllnfvfl-fnivr 288

QY 17 ESGGLLHEIFTSPNLNLLGLGICIFLYKIVR 47

RESULT 8  
ID R90574 standard; Protein; 411 AA.

R90574;  
DE 08-APR-1996 (first entry)  
DE Rat CRF2-alpha receptor.  
KW CRF2-alpha receptor; corticotropin-releasing factor-2 receptor;  
KW cerebrovascular disorder; memory disorder; Alzheimer disease.  
OS Rattus sp.  
FH Key Location/Qualifiers  
FT Domain 1..117  
FT /label= Extracellular\_N-terminal\_domain  
FT Domain 118..138  
FT /label= Transmembrane\_domain  
FT Domain 139..147  
FT /label= Intracellular\_domain  
FT Domain 148..167  
FT /label= Transmembrane\_domain  
FT Domain 168..184  
FT /label= Extracellular\_domain  
FT Domain 185..208  
FT /label= Transmembrane\_domain  
FT Domain 209..223  
FT /label= Intracellular\_domain  
FT Domain 224..244  
FT /label= Transmembrane\_domain  
FT Domain 245..261  
FT /label= Extracellular\_domain  
FT Domain 262..286

FT /label= Transmembrane\_domain  
FT Domain 287..309  
FT /label= Intracellular\_domain  
FT Domain 310..329  
FT /label= Transmembrane\_domain  
FT Domain 330..342  
FT /label= Extracellular\_domain  
FT Domain 343..363  
FT /label= Transmembrane\_domain  
FT Domain 364..411  
FT /label= C-terminal\_intracellular\_domain  
PN W09534651-A2.  
PD 21-DEC-1995.  
PF 14-JUN-1995; U07757.  
PR 14-JUN-1994; US-259959.  
PR 31-JAN-1995; US-381433.  
PR 07-JUN-1995; US-485984.  
PR (NEUR-) NEUROCRINE BIOSCIENCES INC.  
PI Chalmers D, De Souza EB, Grigoriadis DE, Liaw CW;  
PI Lovenberg TW, Oltersdorf T;  
PI WPI; 96-049680/05.  
DR N-PSDB; T12243.  
PT Corticotropin-releasing factor-2 receptor, and DNA encoding it -  
PT used to isolate CRF-2 receptor antagonists for the treatment of  
PT cerebrovascular disorders, memory disorders and Alzheimer's disease  
PS Claim 13; Page 70-73; 109pp; English.  
CC Rat corticotropin-releasing factor-2-alpha (CRF2-alpha) receptor  
CC (R90574) is a membrane-bound G-coupled protein receptor involved  
CC in signal transduction. It can be produced by expression of  
CC encoding cDNA (T12243) in prokaryotic or eucaryotic host cells.  
CC Recombinant CRF2 receptor is used to screen CRF2 receptor agonists  
CC and antagonists of therapeutic appln., and to prepare antibodies  
CC which specifically bind to CRF2 receptors.  
SQ Sequence 411 AA;

Query Match 5.8%; Score 89; DB 15; Length 411;

Best Local Similarity 41.9%; Pred. No. 2.50e+01;

Matches 13; Conservative 9; Mismatches 8; Indels 1; Gaps 1;

Db 259 epgdldvdygqpllvlllnfvfl-fnivr 288

QY 17 ESGGLLHEIFTSPNLNLLGLGICIFLYKIVR 47

RESULT 9

ID R90576 standard; Protein; 411 AA.

R90576;  
DE 08-APR-1996 (first entry)  
DE Human CRF2 receptor.  
KW CRF2 receptor; corticotropin-releasing factor-2 receptor;  
KW cerebrovascular disorder; memory disorder; Alzheimer disease.  
OS Homo sapiens.  
PN W09534651-A2.  
PD 21-DEC-1995.  
PF 14-JUN-1995; U07757.  
PR 14-JUN-1994; US-259959.  
PR 31-JAN-1995; US-381433.  
PR 07-JUN-1995; US-485984.  
PR (NEUR-) NEUROCRINE BIOSCIENCES INC.  
PI Chalmers D, De Souza EB, Grigoriadis DE, Liaw CW;  
PI Lovenberg TW, Oltersdorf T;  
PI WPI; 96-049680/05.  
DR N-PSDB; T12247.  
PT Corticotropin-releasing factor-2 receptor, and DNA encoding it -  
PT used to isolate CRF-2 receptor antagonists for the treatment of  
PT cerebrovascular disorders, memory disorders and Alzheimer's disease  
PS Disclosure; Page 80-82; 109pp; English.  
CC Human corticotropin-releasing factor-2 (CRF2) receptor (R90576) is a  
CC membrane-bound G-coupled protein receptor involved in signal  
CC transduction. It can be produced by expression of encoding cDNA  
CC (T12247) in prokaryotic or eucaryotic host cells. Recombinant CRF2  
CC receptor is used to screen CRF2 receptor agonists and antagonists of  
CC therapeutic appln., and to prepare antibodies which specifically bind





07-JUN-1995; US-485984.  
(NEUR-) NEUROCRINE BIOSCIENCES INC.  
PI Chalmers D, De Souza EB, Grigoriadis DE, Liaw CW;  
PI Lovenberg TW, Oltersdorf T;  
PI WPI; 96-049680/05.  
N-PSDB; TI2244.  
DR Corticotropin-releasing factor-2 receptor, and DNA encoding it -  
PT used to isolate CRF2 receptor antagonists for the treatment of  
PT cerebrovascular disorders, memory disorders and Alzheimer's disease  
PS Disclosure; page 63-66; 109pp; English.  
CC Rat corticotropin-releasing factor-2-beta (CRF2-beta) receptor  
CC (R90575) is a membrane-bound G-coupled protein receptor involved  
CC in signal transduction. It can be produced by expression of  
CC encoding cDNA (TI2244) in procaryotic or eucaryotic host cells.  
CC Recombinant CRF2 receptor is used to screen CRF2 receptor agonists  
CC and antagonists of therapeutic appln., and to prepare antibodies  
CC which specifically bind to CRF2 receptors.  
SQ Sequence 431 AA;

Query Match 5.8%; Score 89; DB 15; Length 431;  
Best Local Similarity 41.9%; Pred. No. 2.50e+01;  
Matches 13; Conservative 9; Mismatches 8; Indels 1; Gaps 1;

Db 279 epgdldvdyggpiilvlvllnfvfl-fnivr 308  
| : | : | : : | : | : | : | : | : | :  
Qv 17 FSGGLHEIFTSPLNLLGLCFLLYKIVR 47

| RESULT | I4                                |
|--------|-----------------------------------|
| ID     | R66796 standard; Protein; 289 AA. |

DE Novel mouse proteoglycan syndecan-1 putative mature protein sequence.  
DI 11-SEP-1999 (first entry)  
DE Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
KW ectodermal; NMGs; glycosylation; heparan sulphate; chondroitin sulphate;  
KW chimaera; chimaeric molecule; effector molecule; receptor; drug;  
KW antibody; diagnostic agent  
KW

|     |   |                     |
|-----|---|---------------------|
| OS  | antibody; diagnostic agent.               |                     |
| WS  | mus musculus.                             |                     |
| KS  | Key                                       | Location/Qualifiers |
| FH  | Domain                                    | 1..230              |
| FT  | /label= extracellular domain              |                     |
| FT  | /label= Domain                            | 23..26              |
| FT  | /note= "heparan sulphate attachment site" |                     |
| FT  | Domain                                    | 231..255            |
| FT  | /label= transmembrane domain              |                     |
| FT  | Domain                                    | 256..289            |
| FT  | /label= intracellular domain              |                     |
| FT  | W09500633-A.                              |                     |
| PN  | 05-JAN-1995.                              |                     |
| PD  | 17-JUN-1994;                              | U06920.             |
| PP  | 17-JUN-1993;                              | US-078683.          |
| PP  | (CHIL-) CHILDRENS MEDICAL CENT.           |                     |
| PPA | (STRD) UNIV LELAND STANFORD JUNIOR.       |                     |
| PI  | Bernfield M, Kato M, Saunders S;          |                     |
| PI  | WPI: 95-052071/07.                        |                     |

DNA and protein sequences for recombinant syndecan-derived proteoglycans - comprising a core protein having glycosylation sites for heparin sulphate glycosaminoglycan side chains. Claim 13; Page 81; 97pp; English.

The sequence of the mature cell surface proteoglycan - syndecan-1. The corresponding gene (Q81748) was isolated from a mouse mammary epithelial cell cDNA library in lambda gtl1 using rabbit antisera against the ectodomain of NMuMG mouse mammary epithelial cell syndecan-1. The complete sequence of the gene encodes a protein(R66793) of 30-35 kD comprising: (1) a hydrophilic N-terminal extracellular domain (residues 23-252); (2) a hydrophobic transmembrane domain (residues 253-277) and (3) a hydrophilic C-terminal intracellular domain (residues 278-331).

The protein contains a protease susceptible cleavage sequence extracellularly and adjacent to the transmembrane region and at least one glycosylation site (residues 45-48) for the attachment of a heparan sulphate chain to the extracellular region. The syndecan-1 protein is thought to contain a 22 amino acid signal peptide, lacking in this sequence, but this was unresolved due to N-terminal blocking of the mature peptide.



## (TM)

```

mpsrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 09:49:41 1997; MasPar time 6.39 Seconds
Tabular output not generated. 730.317 Million cell updates/sec

```

Scoring table: PAM 150  
Gap 11

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Statistics: Mean 46.631; Variance 82.019; scale 0.569

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.               |          |
|------------|-------|-------------|--------|-------|-------------|-------------------------|----------|
| 1          | 119   | 7.7         | 379    | 7     | PDVS_NPVCF  | PUTATIVE PDV-SPECIFIC   | 2.34e-04 |
| 2          | 103   | 6.7         | 375    | 7     | PDVS_NPVCF  | PUTATIVE PDV-SPECIFIC   | 5.75e-02 |
| 3          | 100   | 6.5         | 327    | 9     | SDS3_YEAST  | SDS3 PROTEIN.           | 1.52e-01 |
| 4          | 97    | 6.3         | 106    | 1     | BNZC_PSEUP  | BENZENE 1,2-DIOXYGENA   | 3.94e-01 |
| 5          | 97    | 6.3         | 492    | 2     | CPBC_RAT    | CYCLOCHROME P450 IIB12  | 3.94e-01 |
| 6          | 95    | 6.2         | 107    | 1     | BDDB_PSEUP  | BENZENE 1,2-DIOXYGENA   | 7.34e-01 |
| 7          | 94    | 6.1         | 116    | 11    | VJ57_YEAST  | HYPOTHETICAL 13.1 KD    | 9.98e-01 |
| 8          | 94    | 6.1         | 242    | 6     | NIFY_AZOVI  | NIFY PROTEIN.           | 9.98e-01 |
| 9          | 93    | 6.0         | 106    | 9     | TODB_PSEUP  | TOLUENE 1,2-DIOXYGENA   | 1.35e+00 |
| 10         | 92    | 6.0         | 237    | 4     | HETI_ANASP  | HETI PROTEIN (FRAGMENT) | 1.83e+00 |
| 11         | 93    | 6.0         | 428    | 6     | MANA_YEAST  | MANNOSE-6-PHOSPHATE I   | 1.35e+00 |
| 12         | 93    | 6.0         | 491    | 2     | CPBL_RAT    | CYCLOCHROME P450 IIB1   | 1.35e+00 |
| 13         | 92    | 6.0         | 522    | 3     | CYDA_ECOLI  | CYCLOCHROME D UBIQUINO  | 1.83e+00 |
| 14         | 92    | 6.0         | 613    | 2     | CIKG_RAT    | POTASSIUM CHANNEL PRO   | 1.83e+00 |
| 15         | 92    | 6.0         | 635    | 2     | CIKF_RAT    | POTASSIUM CHANNEL PRO   | 1.83e+00 |
| 16         | 92    | 6.0         | 638    | 2     | CIKE_RAT    | POTASSIUM CHANNEL PRO   | 1.83e+00 |
| 17         | 91    | 5.9         | 349    | 7     | PHOE_KLEOX  | OUTER MEMBRANE PORE P   | 2.47e+00 |
| 18         | 91    | 5.9         | 491    | 2     | CPB2_RAT    | CYCLOCHROME P450 IIB2   | 2.47e+00 |
| 19         | 90    | 5.8         | 64     | 3     | KROS_DROME  | DROSOCIN PRECURSOR.     | 3.32e+00 |
| 20         | 90    | 5.8         | 327    | 8     | RPPV_PIRVY  | RNA POLYMERASE ALPHA    | 3.32e+00 |
| 21         | 89    | 5.8         | 411    | 2     | CFR2_RAT    | CORTICOTROPIN RELEASE   | 4.45e+00 |
| 22         | 89    | 5.8         | 836    | 7     | PAPC_ECOLI  | OUTER MEMBRANE USHER    | 4.45e+00 |



KA MEDLINE; 90431970.

CC : SUBUNIT TWO

CC FERREDOXIN (TODD) AND A FERREDOXIN REDUCTASE (TODA).  
CC -!- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE  
CC FERREDOXIN COMPONENTS.

DR EMBL; J04996; G151603; -.  
DR PIR; C36516; C36516.

KW AROMATIC HYDROCARBONS CATABOLISM; ELECTRON TRANSPORT; IRON-SULFUR.

FT INIT\_MET 0 0  
FT METAL 42 42 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 44 44 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 61 61 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 64 64 IRON-SULFUR (2FE-2S) (POTENTIAL).  
SQ SEQUENCE 106 AA; 11758 MW; E2CC381A CRC32;

Query Match 6.0%; Score 93; DB 9; Length 106;

Best Local Similarity 30.3%; Pred. No. 1.35e+00;

Matches 10; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

Db 6 lrrgdppemryegpvmvncvdgeffav 38

QY 68 LKRRDTPAELRRFDGVPRLMAINGKVFV 100

RESULT 10

ID HETLANASP STANDARD; PRT; 237 AA.

AC P37695;

DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)

DE HETI PROTEIN (FRAGMENT).

GN HETI.

OS ANABAENA SP. (STRAIN PCC 7120).

OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;

OC CYANOBACTERIA (BLUE-GREEN ALGAE); NOSTOCALES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 94209228.

RA BLACK T.A., WOLK C.P.;

RL J. BACTERIOL. 176:2282-2292(1994).

CC -!- FUNCTION: MAY BE REQUIRED FOR MAINTAINING VEGETATIVE GROWTH AND

CC -!- PROBABLY ACTS VIA HETN TO INHIBIT DIFFERENTIATION.

CC -!- SIMILARITY: BELONGS TO THE ENTD/GSP/HETI/SFP FAMILY.

DR EMBL; L22883; G441121; ALT\_INIT.

FT NON\_TER 1 1

SQ SEQUENCE 237 AA; 27090 MW; CB40CD18 CRC32;

Query Match 6.0%; Score 92; DB 4; Length 237;

Best Local Similarity 32.7%; Pred. No. 1.83e+00;

Matches 36; Conservative 21; Mismatches 42; Indels 11; Gaps 10;

Db 1 llightwlpkpnltllsdevhl-wri-pldqpsqlqdlaatlssdelaranfyf-peh 57

QY 21 LLHEI-FTSPLNLLGLCFLYKIVRGDPAAS-GDRTTXPPP-LPRLKRDETPAE 77

Db 58 rrrftagrg--llrsilgylgvepqvkvfdesrgk-pllgdrfaesgi 104

QY 78 LRRFDGVPRLMAINGKVFDTKGR-KF-YGPEGYGVFAGRDASRGL 125

RESULT 11

ID MANA\_YEAST

AC P29952.

DT 01-APR-1993 (REL. 25, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE MANNOSE-6-PHOSPHATE ISOMERASE (EC 5.3.1.8) (PHOSPHOMANNOSE ISOMERASE)

DE (PMT) (PHOSPHOHEXOMUTASE).

GN PMT40 OR YER003C.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE: 92318908.

RA SMITH D.J., PROUDFOOT A.E.I., FRIEDLI L., KLIG L.S., PARAVICINI G.,

RA PAYTON M.A.;  
RL MOL. CELL. BIOL. 12:2924-2930(1992).  
RN [2]  
RC STRAIN-S288C / AB972;  
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,  
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,  
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICK-SMITH S.,  
RA HYMAN R., RAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,  
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,  
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,  
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;  
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RN INHIBITION BY ZINC.  
RX MEDLINE: 93192232.  
RA WELLS T.N.C., COULIN F., PAYTON M.A., PROUDFOOT A.E.I.;  
RL BIOCHEMISTRY 32:1294-1301(1993).  
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE GDP-MANNOSE AND  
CC DOLICHOL-PHOSPHATE-MANNOSE REQUIRED FOR A NUMBER OF CRITICAL  
CC MANNOsyl TRANSFER REACTIONS.  
CC -!- CATALYTIC ACTIVITY: D-MANNOSE 6-PHOSPHATE = D-FRUCTOSE  
CC 6-PHOSPHATE.  
CC -!- COFACTOR: ZINC.  
CC -!- SUBUNIT: MONOMER.  
CC -!- ENZYME REGULATION: CAN BE INHIBITED BY AN EXCESS OF ZINC.  
CC -!- PATHWAY: GLYCOSYLATION; EARLY STEPS OF MANNOsylATION.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- INDUCTION: BY D-MANNOSE.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF MANNOSE-6-PHOSPHATE ISOMERASES.  
DR EMBL; M85238; G172166; -.  
DR EMBL; U18778; G603595; -.  
DR PIR; S31240; S31240.  
DR LISTA; SC00834; PM140.  
DR SGD; L0001452; PM140.  
DR PROSITE; PS00965; PM1\_I\_1.  
DR PROSITE; PS00966; PM1\_I\_2.  
KW ISOMERASE; ZINC; ACETYLATION.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT METAL 108 108 ZINC (BY SIMILARITY).  
FT METAL 110 110 ZINC (BY SIMILARITY).  
FT METAL 135 135 ZINC (BY SIMILARITY).  
FT METAL 280 280 ZINC (BY SIMILARITY).  
FT CONFLICT 24 24 A -> R (IN REF. 1).  
SQ SEQUENCE 428 AA; 48057 MW; C5B84A72 CRC32;

Query Match 6.0%; Score 93; DB 6; Length 428;

Best Local Similarity 26.5%; Pred. No. 1.35e+00;

Matches 13; Conservative 16; Mismatches 19; Indels 1; Gaps 1;

Db 366 rhfegvdgpsillttkngyikadgkkl-kaepgfvffiahlipvdlea 413

QY 79 RRDGVQDPRLMAINGKVFDTKGRKFGYGPYGVFAGRDASRGLAT 127

RESULT 12

ID CPB1\_RAT

AC P00176;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE CYTOCHROME P450 11B1 (EC 1.14.14.1) (P450-B) (P450-PB1 AND P450-PB2)

DE (P450-LM2) (PHENOBARBITAL-INDUCIBLE).

GN CYP2B-1.

OS RATRUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE OF 6-491 FROM N.A. (ISOZYME PB1 AND PB2).

RX MEDLINE: 82222224.

RA FUJII-KURIYAMA Y., MIZUKAMI Y., KAWAJIRI K., SOGAWA K., MURAMATSU M.;

RL PROC. NATL. ACAD. SCI. U.S.A. 79:2793-2797(1982).

```

RN ISOZYMES PB1 AND PB2, REVISIONS.
RA FUJIT-KURIYAMA Y., MIZUKAMI Y., KAWAJIRI K., SOGAWA K., MURAMATSU M.;
RL PROC. NATL. ACAD. SCI. U.S.A. 79:5443-5443(1982).
RN
RN
RN SEQUENCE OF 1-22.
RX MEDLINE; 79194111.
RA BOTELHO L.H., RYAN D.E., LEVIN W.;
RL J. BIOL. CHEM. 254:5635-5640(1979).
RN
RN PHOSPHORYLATION.
RX MEDLINE; 90059885.
RA PYERIN W., TANIGUCHI H.;
RL EMBO J. 8:3003-3010(1989).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES IN LIVER MICROSOMES. THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- INDUCTION: BY PHENOBARBITAL.
CC -1- PFM: PHOSPHORYLATION IS ACCOMPANIED BY A DECREASE IN ENZYME
CC ACTIVITY.
DR EMBL; J00719; G203753; -.
DR EMBL; M37134; G203785; -.
DR PIR; A00176; O4RTPB.
DR PROSITE; PS00086; CYTOCHROME_P450.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME;
KW MICROsome; PHOSPHORYLATION.
FT MOD_RES 128 128 PHOSPHORYLATION (BY CAPK).
FT BINDING 436 436 HEME.
FT VARIANT 303 303 S -> G (IN ISOZYME PB2).
FT VARIANT 321 322 AE -> TV (IN ISOZYME PB2).
FT VARIANT 337 337 L -> P (IN ISOZYME PB2).
FT VARIANT 339 339 T -> S (IN ISOZYME PB2).
FT VARIANT 344 344 S -> T (IN ISOZYME PB2).
SQ SEQUENCE 491 AA; 55933 MW; E6F42ED3 CRC32;

Query Match 6.0%; Score 93; DB 2; Length 491;
Best Local Similarity 46.3%; Pred. No. 1.35e+00;
Matches 19; Conservative 6; Mismatches 12; Indels 4; Gaps 4;

Db 3 ptilllallvgflll-lvrg-hpksrgn-fppgprlppli 40
29 PLNLLLLGLCI-FLLYKIVRGDPAASGDRTTTTPPLPRL 68

RESULT 13
ID CYDA_ECOLI STANDARD; PRT; 522 AA.
AC P11026;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE CYTOCHROME D UBIQUINOL OXIDASE SUBUNIT I (EC 1.10.3.-).
GN CYDA OR CYD-1.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN
RN
RN SEQUENCE FROM N.A.
RX MEDLINE; 88330812.
RA GREEN G.N., FANG H., LIN R.-J., NEWTON G., MATHER M., GEORGIU C.D.,
RA GENNIS R.B.;
RL J. BIOL. CHEM. 263:13138-13143(1988).
RN
RN
RN TOPOLOGY.
RX MEDLINE; 88330811.
RA GEORGIU C.D., DUWEKE T.J., GENNIS R.B.;
RL J. BIOL. CHEM. 263:13130-13137(1988).
RN
RN
RN DOMAINS.

RX MEDLINE; 90170924.
RA DUWEKE T.J., GENNIS R.B.;
RL J. BIOL. CHEM. 265:4273-4277(1990).
CC -1- FUNCTION: CYTOCHROME D TERMINAL OXIDASE COMPLEX IS THE COMPONENT
CC OF THE AEROBIC RESPIRATORY CHAIN OF E. COLI THAT PREDOMINATES WHEN
CC CELLS ARE GROWN AT LOW AERATION.
CC -1- CATALYTIC ACTIVITY: UBIQUINOL-8 + O(2) -> UBIQUINONE-8 + H(2)O.
CC -1- COFACTOR: CONTAINS THE PROTHIEME IX CENTER B558.
CC -1- SUBUNIT: HETERODIMER OF SUBUNITS I AND II.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: STRONG, TO E. COLI APPC.
DR EMBL; J03939; G497637; -.
DR PIR; A28940; A28940.
DR EMBL; J03939; G497637; -.
KW OXIDOREDUCTASE; ELECTRON TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE;
KW HEME; FORMYLATION.
FT MOD_RES 1 1 FORMYLATION.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 42 POTENTIAL.
FT DOMAIN 43 94 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 95 114 POTENTIAL.
FT DOMAIN 115 129 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 130 149 POTENTIAL.
FT DOMAIN 150 187 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 188 207 POTENTIAL.
FT DOMAIN 208 219 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 220 239 POTENTIAL.
FT DOMAIN 240 392 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 393 412 POTENTIAL.
FT DOMAIN 413 470 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 471 490 POTENTIAL.
FT DOMAIN 491 522 PERIPLASMIC (POTENTIAL).
FT METAL 186 186 IRON (HEME B558 AXIAL LIGAND).
SQ SEQUENCE 522 AA; 58171 MW; 0FB87D01 CRC32;

Query Match 6.0%; Score 92; DB 3; Length 522;
Best Local Similarity 26.3%; Pred. No. 1.83e+00;
Matches 15; Conservative 15; Mismatches 26; Indels 1; Gaps 1;

Db 452 aigevlptavansltagdlifsmvlglyclflvaelfmfkfarlgpslskgr 508
2 AAEDVATGADPSDLESGLLHE-IFTSPLNLLGLCFLLYKIVRGDPAASGDR 57

RESULT 14
ID CIXG_RAT STANDARD; PRT; 613 AA.
AC P22463;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE POTASSIUM CHANNEL PROTEIN KV3.2 (KSHIIIA.1).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN
RN
RN SEQUENCE FROM N.A.
RX TISSUE-BRAIN;
RX MEDLINE; 90311375.
RA MCCORMACK T., DE MIERA E.C.V.-S., RUDY B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:5227-5231(1990).
RN
RN
RN REVISIONS.
RX MEDLINE; 91219512.
RA MCCORMACK T., DE MIERA E.C.V.-S., RUDY B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:4060-4060(1991).
CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM
CC ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED
CC CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE
CC MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH
CC WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL
CC GRADIENT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS CHARACTERIZED

```

CC BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD  
CC POSITION.  
CC -1- THE TAIL MAY BE IMPORTANT IN MODULATION OF CHANNEL ACTIVITY AND/OR  
CC TARGETING OF THE CHANNEL TO SPECIFIC SUBCELLULAR COMPARTMENTS.  
CC -1- THIS CHANNEL PROTEIN BELONGS TO THE DELAYED RECTIFIER CLASS.  
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT FORMS OF KV3.2 ARE PRODUCED  
CC BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: BELONGS TO SHAW POTASSIUM CHANNEL SUBFAMILY.  
DR EMBL; M34052; G206914; -  
KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;  
KW GLYCOPROTEIN; MULTIGENE FAMILY; ALTERNATIVE SPLICING; PHOSPHORYLATION.  
FT DOMAIN 1 229 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 230 248 SEGMENT S1 (POTENTIAL).  
FT DOMAIN 249 283 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 284 303 SEGMENT S2 (POTENTIAL).  
FT DOMAIN 304 314 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 315 337 SEGMENT S3 (POTENTIAL).  
FT DOMAIN 338 345 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 346 368 SEGMENT S4 (POTENTIAL).  
FT DOMAIN 369 381 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 382 401 SEGMENT S5 (POTENTIAL).  
FT DOMAIN 402 450 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 451 473 SEGMENT S6 (POTENTIAL).  
FT DOMAIN 474 613 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 56 99 GLY/PRO-RICH (INSERT).  
FT CARBOHYD 259 259 POTENTIAL.  
FT CARBOHYD 266 266 POTENTIAL.  
FT MOD\_RES 564 564 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT MOD\_RES 596 596 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
FT MOD\_RES 606 606 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
FT MOD\_RES 610 610 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
SQ SEQUENCE 613 AA; 67550 MW; 8A92FB2C CRC32;  
Query Match 6.0%; Score 92; DB 2; Length 613;  
Best Local Similarity 32.7%; Pred. No. 1.83e+00;  
Matches 34; Conservative 14; Mismatches 51; Indels 5; Gaps 5;  
Db 12 Invgttrhetyrstkltp-gtrlallassepggdcltaagdklqlppplspppprpppl 70  
QY 16 LESGGLLHEIFTSPLNLLGLICIFLL-YKIVRGDQPAASGDRTTXPPPL-PLKRRDF 73  
Db 71 spvpsgcfcgagncsshgngsd-hpgggreffdrhp-gvfa 112  
QY 74 TPAELRRFDGVQDPRILMAINGKVFDTGKRKFYGPYGVFA 117  
RESULT 15  
ID CIXF\_RAT STANDARD; PRT; 635 AA.  
AC P22461;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
DE POTASSIUM CHANNEL PROTEIN KV3.2C.  
OS RATUUS NORVEGICUS (RAT).  
NC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
AC EUTHERIA; RODENTIA.  
HN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 91348257.  
RA LUNEAU C.J., WIEDMANN R., SMITH J.S., WILLIAMS J.B.;  
RL FEBS LETT. 288:163-167(1991).  
CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM  
CC ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED  
CC CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE  
CC MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH  
CC WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL  
CC GRADIENT.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS CHARACTERIZED  
CC BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD  
CC POSITION.  
CC -1- THE TAIL MAY BE IMPORTANT IN MODULATION OF CHANNEL ACTIVITY AND/OR  
CC TARGETING OF THE CHANNEL TO SPECIFIC SUBCELLULAR COMPARTMENTS.  
CC -1- THIS CHANNEL PROTEIN BELONGS TO THE DELAYED RECTIFIER CLASS.  
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT FORMS OF KV3.2 ARE PRODUCED  
CC BY ALTERNATIVE SPLICING OF THE SAME GENE.

CC -1- SIMILARITY: BELONGS TO SHAW POTASSIUM CHANNEL SUBFAMILY.  
DR EMBL; M59313; G206046; ALT\_SEQ.  
KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;  
KW GLYCOPROTEIN; MULTIGENE FAMILY; ALTERNATIVE SPLICING; PHOSPHORYLATION.  
FT DOMAIN 1 229 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 230 248 SEGMENT S1 (POTENTIAL).  
FT DOMAIN 249 283 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 284 303 SEGMENT S2 (POTENTIAL).  
FT DOMAIN 304 314 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 315 337 SEGMENT S3 (POTENTIAL).  
FT DOMAIN 338 345 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 346 368 SEGMENT S4 (POTENTIAL).  
FT DOMAIN 369 381 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 382 401 SEGMENT S5 (POTENTIAL).  
FT DOMAIN 402 450 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 451 473 SEGMENT S6 (POTENTIAL).  
FT DOMAIN 474 635 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 56 99 GLY/PRO-RICH (INSERT).  
FT CARBOHYD 259 259 POTENTIAL.  
FT CARBOHYD 266 266 POTENTIAL.  
FT MOD\_RES 564 564 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT MOD\_RES 596 596 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
FT MOD\_RES 606 606 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
FT MOD\_RES 610 610 PHOSPHORYLATION (BY CK-II) (POTENTIAL).  
SQ SEQUENCE 635 AA; 69859 MW; 703505AD CRC32;  
Query Match 6.0%; Score 92; DB 2; Length 635;  
Best Local Similarity 32.7%; Pred. No. 1.83e+00;  
Matches 34; Conservative 14; Mismatches 51; Indels 5; Gaps 5;  
Db 12 Invgttrhetyrstkltp-gtrlallassepggdcltaagdklqlppplspppprpppl 70  
QY 16 LESGGLLHEIFTSPLNLLGLICIFLL-YKIVRGDQPAASGDRTTXPPPL-PLKRRDF 73  
Db 71 spvpsgcfcgagncsshgngsd-hpgggreffdrhp-gvfa 112  
QY 74 TPAELRRFDGVQDPRILMAINGKVFDTGKRKFYGPYGVFA 117

Search completed: Thu Nov 6 09:50:11 1997  
Job time : 30 secs.



\*\*\*\*\*

MPERCH\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Release 2.1D John F. Collins, Biocomputing Research Unith.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

Run on: Fri Nov 7 15:14:48 1997; MasPar time 212.87 Seconds  
Tabular output not generated.  
1017.291 Million cell updates/sec

Title: >US-08-822-264-2  
Description: (1-788) from US08822264.seq  
Perfect Score: 784  
N.A. Sequence: 1 GCCGCGCGTGGCGCGCGG.....AATATTAGAAAGTTTGAGC 788  
Comp: CGCGCGCGTGGCGCGCGG.....TTATAATCTTCAACTCG

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 359085 seqs, 137405154 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13  
EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnSTS 75:enEST1 76:enEST2 77:enEST3 78:enEST4  
79:enEST5 80:enEST6 81:enEST7 82:enEST8 83:enEST9  
84:enEST10 85:enEST11 86:enEST12 87:enEST13 88:enEST14  
89:enEST15 90:enEST16 91:enEST17 92:enEST18 93:enEST19  
94:enEST20 95:enEST21 96:enEST22 97:enEST23 98:enEST24  
99:enEST25 100:enEST26 101:enEST27 102:enEST28  
103:enEST29 104:enEST30 105:enEST31 106:enSTS 107:ueEST1  
108:ueEST2

Database: EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnSTS 75:enEST1 76:enEST2 77:enEST3 78:enEST4  
79:enEST5 80:enEST6 81:enEST7 82:enEST8 83:enEST9  
84:enEST10 85:enEST11 86:enEST12 87:enEST13 88:enEST14  
89:enEST15 90:enEST16 91:enEST17 92:enEST18 93:enEST19  
94:enEST20 95:enEST21 96:enEST22 97:enEST23 98:enEST24  
99:enEST25 100:enEST26 101:enEST27 102:enEST28  
103:enEST29 104:enEST30 105:enEST31 106:enSTS 107:ueEST1  
108:ueEST2

Statistics: Mean 10.814; Variance 2.079; scale 5.201  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result Query  
SUMMARIES

| No. | Score | Match | Length | DB  | ID        | Description           | Pred. No. |
|-----|-------|-------|--------|-----|-----------|-----------------------|-----------|
| 1   | 402   | 51.3  | 535    | 104 | MMAA84910 | mt62c03.r1 Soares 2Nb | 0.00e+00  |
| 2   | 402   | 51.3  | 535    | 61  | AA184910  | mt62c03.r1 Soares 2Nb | 0.00e+00  |
| 3   | 402   | 51.3  | 535    | 13  | AA184910  | mt62c03.r1 Soares 2Nb | 0.00e+00  |
| 4   | 400   | 51.0  | 458    | 24  | N52291    | Yv48a12.s1 Soares fet | 0.00e+00  |
| 5   | 387   | 49.4  | 501    | 28  | AA023630  | mh80f05.r1 Soares mou | 0.00e+00  |
| 6   | 380   | 48.5  | 512    | 5   | AA117455  | mh28g11.r1 Beddington | 0.00e+00  |
| 7   | 377   | 48.1  | 429    | 3   | AA101294  | zn71f03.r1 Stratagene | 0.00e+00  |
| 8   | 352   | 44.9  | 476    | 3   | AA111285  | mo53h02.r1 Life Tech  | 0.00e+00  |
| 9   | 330   | 42.1  | 406    | 4   | AA115422  | zl85e09.r1 Stratagene | 0.00e+00  |
| 10  | 309   | 39.4  | 438    | 21  | AA021062  | ze67a12.r1 Soares ret | 0.00e+00  |
| 11  | 297   | 37.9  | 349    | 65  | AA232394  | zr27f02.r1 Stratagene | 0.00e+00  |
| 12  | 297   | 37.9  | 349    | 82  | HS1150107 | zr27f02.r1 Stratagene | 0.00e+00  |
| 13  | 289   | 36.9  | 344    | 25  | N58287    | Yv68c06.s1 Soares fet | 0.00e+00  |
| 14  | 280   | 35.7  | 446    | 26  | SZ884149  | S.scrofa mRNA; expres | 0.00e+00  |
| 15  | 279   | 35.6  | 381    | 55  | AA088127  | mn89a10.r1 Stratagene | 0.00e+00  |
| 16  | 268   | 34.2  | 352    | 33  | AA106719  | mm19h05.r1 Stratagene | 0.00e+00  |
| 17  | 268   | 34.2  | 359    | 95  | NM4857    | ma55h11.r1 Soares mou | 0.00e+00  |
| 18  | 208   | 26.5  | 215    | 4   | AA116041  | zm79b10.r1 Stratagene | 0.00e+00  |
| 19  | 170   | 21.7  | 506    | 62  | AA224015  | zr13d05.r1 Stratagene | 1.35e-289 |
| 20  | 170   | 21.7  | 506    | 81  | HS1143543 | zr13d05.r1 Stratagene | 1.35e-289 |
| 21  | 130   | 16.6  | 461    | 12  | AA179573  | zp48f01.r1 Stratagene | 1.14e-205 |
| 22  | 122   | 15.6  | 436    | 29  | AA047623  | zr13b05.r1 Soares fet | 4.04e-189 |
| 23  | 101   | 12.9  | 139    | 33  | AA106610  | mm17h01.r1 Stratagene | 3.31e-146 |
| 24  | 100   | 12.8  | 449    | 96  | NM8175    | ma48e09.r1 Soares mou | 3.47e-144 |
| 25  | 80    | 10.2  | 144    | 25  | N55620    | ESTG177 Rat pancreati | 2.38e-104 |
| 26  | 65    | 8.3   | 480    | 89  | HSW4936   | ze43d08.r1 Soares ret | 1.97e-75  |
| 27  | 65    | 8.3   | 480    | 2   | W96493    | ze43d08.r1 Soares ret | 1.97e-75  |
| 28  | 34    | 4.3   | 179    | 65  | AA231657  | RZ583.F CDNA from ric | 7.14e-21  |
| 29  | 34    | 4.3   | 179    | 105 | OSAA31657 | RZ583.F CDNA from ric | 7.14e-21  |
| 30  | 30    | 3.8   | 301    | 77  | AT66019   | 20700 Arabidopsis tha | 8.18e-15  |
| 31  | 29    | 3.7   | 535    | 26  | N75082    | Yz29a02.r1 Soares mul | 2.32e-13  |
| 32  | 29    | 3.7   | 612    | 79  | ATA41129  | 24395 CD4-13 Arabidop | 2.32e-13  |
| 33  | 27    | 3.4   | 156    | 36  | DM36D7S   | D.melanogaster Sts d  | 1.54e-10  |
| 34  | 27    | 3.4   | 308    | 13  | AA182915  | zp36a08.r1 Stratagene | 1.54e-10  |
| 35  | 27    | 3.4   | 308    | 80  | HS1142468 | zp36a08.r1 Stratagene | 1.54e-10  |
| 36  | 26    | 3.3   | 618    | 5   | AA117437  | mn21h12.r1 Beddington | 3.56e-09  |
| 37  | 25    | 3.2   | 455    | 75  | AT1313    | 5394 Arabidopsis tha  | 7.61e-08  |
| 38  | 25    | 3.2   | 544    | 78  | AT90214   | 19129 Arabidopsis tha | 7.61e-08  |
| 39  | 24    | 3.1   | 142    | 92  | MM161028  | mw29e01.r1 Soares mou | 1.49e-06  |
| 40  | 24    | 3.1   | 142    | 68  | AA245258  | mw29e01.r1 Soares mou | 1.49e-06  |
| 41  | 24    | 3.1   | 319    | 75  | AT123     | 73 Arabidopsis thalia | 1.49e-06  |
| 42  | 23    | 2.9   | 613    | 79  | ATA42635  | 24867 CD4-13 Arabidop | 2.67e-05  |
| 43  | 22    | 2.8   | 322    | 30  | AA059303  | zf65g05.r1 Soares ret | 4.31e-04  |
| 44  | 22    | 2.8   | 497    | 15  | AA191500  | zp88e09.s1 Stratagene | 4.31e-04  |
| 45  | 22    | 2.8   | 532    | 3   | AA081099  | zn34e04.s1 Stratagene | 4.31e-04  |

ALIGNMENTS

RESULT 1  
ID MMAA84910 standard; RNA; EST; 535 BP.  
AC AA184910;  
NI g1768619  
DT 21-FEB-1997 (Rel. 51, Created)  
DT 22-FEB-1997 (Rel. 51, Last updated, Version 2)  
DE mt62c03.r1 Soares 2NbMT Mus musculus CDNA clone 634468 5' similar  
DE to WP:K07E3.6 CE04722 TRANSLOCATING ATPASE ;  
KW EST.  
OS Mus musculus (house mouse)  
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
OC Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
OC Mus.  
RN [1]  
RP 1-535  
RA Marra M., Hallier L., Allen M., Bowles M., Dietrich N.,  
RA Dubuque T., Geisel S., Kucaba T., Lacy M., Martin J.,  
RA Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,  
RA Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,  
RA Waterston R.;  
RT "The WashU-HMI Mouse EST Project";  
RL Unpublished.





```

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 826      Std Error: 0.00
Seq primer: ml3 -40 forward
High quality sequence stop: 278.

Location/Qualifiers
1...458
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pTT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer;
[5' ACTGCGAAGATTAATAAGACATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTT73 vector. Library
went through one round of normalization. Library.
constructed by Bento Soares and M.Fatima Bonaído."
/clone="245950"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>458)
102 a 126 c 114 q 114 t 2 others

```

| mrna                  | BASE COUNT | 102 a   | 126 c         | 114 g       | 114 t   | 2 others |  |
|-----------------------|------------|---|---------------|-------------|---------|----------|--|
| complement(<1..>458)  |            |   |               |             |         |          |  |
| Query Match           | 51.0%;     | Score 400;  | DB 24;        | Length 458; |         |          |  |
| Best Local Similarity | 96.9%;     | Pred. No. 0.00e+00;   |               |             |         |          |  |
| Matches               | 445;       | Conservative 0;   | Mismatches 7; | Indels 7;   | Gaps 7; |          |  |
| Db                    | 1          | aagacagagtggaagtgttacaaatgatttggcaaaatatacaaaaatagatatatacttcacat | 60            |             |         |          |  |
| Cp                    | 752        | AAGACANAGTGGAGCTGTACAAATGATTTTGCAAAATACANNA-TAGATATACTTCCACT      | 694           |             |         |          |  |
| Db                    | 61         | gaatgctttaatcatcttttccgggaactctcatcttttggttcttctcatctgaatata      | 120           |             |         |          |  |
| Cp                    | 693        | GAATGCTTTAA-CATTTTTCGGGAACCTCATCTTTTGGTTCTTCTCATCTGAGTACA         | 635           |             |         |          |  |
| Db                    | 121        | caatgggctctccccctcttcagcagtttggccacgtgatatacttgaagtgaaact         | 180           |             |         |          |  |
| Cp                    | 634        | CAGTGGGCTCCTCCCCCTCTCAGCAGTTTGGCCAGTGATGATACTTGAAGTGAAC           | 575           |             |         |          |  |
| Db                    | 181        | gagactcccgactcacagagtcctcctgctgggcagcagtgaggtcagaaaaggtcatcgt     | 240           |             |         |          |  |
| Cp                    | 574        | GAGACTCCCACTCAGAGTCTCCTGCTGGCAGCAGTGAGGTGAGAAAGGTCACTCGT          | 515           |             |         |          |  |
| Db                    | 241        | actcatcctcagtgcttccttataccaggcaaaatggtgcaaggccc-tggatgcattctc     | 299           |             |         |          |  |
| p                     | 514        | ACTCATCTTTCACTGCTTCTTTATCCAGGCAAAATGTGGCAAGGCCCTGGATGCATCTC       | 455           |             |         |          |  |
| Db                    | 300        | ttccagcaaaagcccatacaggcccctcngccggtagaattntgggccttttggtcac        | 359           |             |         |          |  |
| Cp                    | 454        | TTCCAGCAAAAGACCCCATACGGGCCCTCGGCCCGGTAGAATT-TGGGGCTTTGGTCA        | 396           |             |         |          |  |
| Db                    | 360        | tcaaacaccttcgcttgatggccatgagtatggcgggggtccttggacgcgcgtcgaaac      | 419           |             |         |          |  |
| Cp                    | 395        | TCGAACACCTTGCCGTGTATGGCCATGAGTATG-CGCGGG-TCCTGGAGCCCGTCGAAC       | 338           |             |         |          |  |
| Db                    | 420        | gcgcgagctgcgggggggtgaagtcgcgcgcgttgagg                            | 458           |             |         |          |  |
| Cp                    | 337        | GCSCGAGCTCGCGGGGG-TGAAGTCGCGCGCTTGAGG                             | 300           |             |         |          |  |

| RESULT     | 5 | AA023630     | 501 bp       | mRNA     | EST       | 21-JAN-1997            |
|------------|---|--------------|--------------|----------|-----------|------------------------|
| LOCUS      |   | mn80f05.r1   | Soares mouse | placenta | 4NbMP13.5 | 14.5 Mus musculus cdNA |
| DEFINITION |   | clone 457281 | 5'           |          |           |                        |
| ACCESSION  |   | AA023630     |              |          |           |                        |
| NID        |   | 91487547     |              |          |           |                        |
| KEYWORDS   |   | EST.         |              |          |           |                        |
| SOURCE     |   | house mouse. |              |          |           |                        |
| ORGANISM   |   | Mus musculus |              |          |           |                        |

Eukaryotae: mitochondrial eukaryotes; Metazoa: Chordata: Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 501)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, G., Geisel, S., Kuback, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE

The WashU-HHMI Mouse EST Project

Unpublished (1996)

JOURNAL

| TITLE  | JOURNAL                         | COMMENT                           |
|--|---------------------------------|-----------------------------------|
| 1. The Role of the Teacher in the Classroom                  | Journal of Educational Research | 1980, Vol. 83, No. 1, pp. 1-10    |
| 2. The Impact of Technology on Education                     | Journal of Educational Research | 1980, Vol. 83, No. 2, pp. 11-20   |
| 3. The Importance of Parental Involvement                    | Journal of Educational Research | 1980, Vol. 83, No. 3, pp. 21-30   |
| 4. The Effect of Teacher Expectations on Student Achievement | Journal of Educational Research | 1980, Vol. 83, No. 4, pp. 31-40   |
| 5. The Role of the School in the Community                   | Journal of Educational Research | 1980, Vol. 83, No. 5, pp. 41-50   |
| 6. The Impact of Teacher Education on Classroom Practice     | Journal of Educational Research | 1980, Vol. 83, No. 6, pp. 51-60   |
| 7. The Importance of Teacher Collaboration                   | Journal of Educational Research | 1980, Vol. 83, No. 7, pp. 61-70   |
| 8. The Effect of Teacher Attitudes on Student Behavior       | Journal of Educational Research | 1980, Vol. 83, No. 8, pp. 71-80   |
| 9. The Role of the School in the Development of the Child    | Journal of Educational Research | 1980, Vol. 83, No. 9, pp. 81-90   |
| 10. The Impact of Teacher Training on Student Learning       | Journal of Educational Research | 1980, Vol. 83, No. 10, pp. 91-100 |

Contact: Karra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.lln.gov) for further information.  
MGI:274169  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 494.

|                 |               |
|-----------------|---------------|
| <b>FEATURES</b> | <b>SOURCE</b> |
|-----------------|---------------|

```

1..501
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
was primed with Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adap
(Pharmacia), digested with Not I and cloned into the N
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."
/clone="457281"
/clone_lib="Soares mouse placenta 4NBMP13.5 14.5"
/sex="unknown"
/dev_stage="adult"
/lab_host="DH10B"
<1..>501
mRNA

```

| mrna     | base count | origin              |
|----------|------------|---------------------|
| <1..>501 | 95 a 159 c | 152 g 94 t 1 others |

Query Match 49.4%; Score 387; DB 28; Length 501;

Best Local Similarity 90.6%; Pred. No. 0.00e+00;

|         |                   |               |            |         |    |
|---------|-------------------|---------------|------------|---------|----|
| Matches | 441; Conservative | 0; Mismatches | 44; Indels | 2; Gaps | 2; |
|---------|-------------------|---------------|------------|---------|----|

Db 16 aagcaggactatccggaqctctgcctagccqgqcccaacctttgctccagaqatcatggc 75

[illegible]

QY 45 AAGTGGCGAGTTCCGGATCCCTGCCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGC 104

[illegible]

Db 76 tggcaggatgtggtggcgactggcgccgagcccgagcgtagaggcgggcggtgct 135

[illegible]

QY TGCCGAGGATGTGGTGGCGACTGGCGCCGACCAAGCGATCTGGAGAGCGGCGGCTGCT 164

**Db** 136 qcacqagattttcacggtctcctctcaacctgctcctccttcttgaggcctctgcacatcttcctgct 195

22 100 99 98 97 96 95 94 93 92 91 90 89 88 87 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67 66 65 64 63 62 61 60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

QY 165 GCATGAGATTTTACAGTCGCCGCTCAACCTGCTGCTGGCCTCTGCATCTTCCTGCT 224

Db 196 ctacaagatcgttcgcggggaccagcccggtgccagtggcgacaacgacgacgaa-c 254

[illegible]

QY 225 CTACAAGATCGTGCCTGGGGGACACGCCGGCGGACAG-GACGACGACGANGC 283

Db 255 caccctggctgccccgcctcaagcggcgcancttcacccctgcccgaactgagcgctttccg 314

[illegible]

Qy 284 CGCCCCCTGCCCCCGCCTCAAGCGGCGGACTTCACCCCCGCCGAGCTGCGGCGCTTCG 343

**F**

Db 315 atggcgtccaggaccgcgcattctcatggccatcaacggcaagggtgttcgacgtgacca 374

244

QY 344 ACGGCGTCCAGGACCGCGCATACTCATGGCCATCAACGGCAAGGTGTTTCGATGTGACCA 403



(Ntera-2/cl.D1). Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' /clone="563645"

/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (Xanadmycin resistant)"  
mRNA  
BASE COUNT 79 a 136 c 131 g 74 t 9 others  
ORIGIN

Query Match 48.1%; Score 377; DB 3; Length 429;  
Best Local Similarity 94.0%; Pred. No. 0.00e+00;  
Matches 40; Conservative 0; Mismatches 24; Indels 2; Gaps 2;

Db 1 agaaagtgccgaggtccggtccctcctagcggcgcccaacctttactccagagatcat 60

Qy 42 AGAAGTGGCGAGTCCGATCCCTGCTTCCGAGCGGCCCAACCTTTACTCCAGAGATCAT 101

Db 61 ggctccgagagatgtggtggcactggcgcnagaccacaaacgacatctggagagcgcggt 120

Qy 102 GGTCCGAGGATGTGGTGGCGACTGGCGCGACCAAGCGATCTGGAGAGCGCGGGCT 161

Db 121 gctgatgagatttcacgctccgctcaacctgctgctgtgctgctcctatgcatcttc 180

Qy 162 GCTGATGAGATTTTACGTCGCCGCTCAACCTGCTGCTGTGGCTC-TGCATCTTCC 220

Db 181 tgcctacaaagatcgtgcggggaccagcgccgcnacgcnagcngancagacgacga 240

Qy 221 TGCTTACAAAGATCTGTCGGGGGACAGCGCGGCCGACGCGGACAGACACGGA 280

Db 241 -gcgcgcctcctgcgcgcctcctcctcctcctcctcctcctcctcctcctcctcct 299

Qy 281 NGCCGCCCTCTGCCCCCTCAAGCGCGGACTTACCCCGCGAGCTGCGCGCT 340

Db 300 tcgaggggtccagaccgacatactatgagccatcaacggcaagggtgttcgagtgga 359

Qy 341 TCGAGCGGCTCCAGGACCGCGCATACTCATGCCCATCAACGGCAAGGTGTCGATGTGA 400

Db 360 caaaagcgcgaattctacggcgccgagcgcnatattgggtcttctggaagagatg 419

Qy 401 CCAAGCGCGCAAAATCTACGGGCGCGGCGGCGGCTATGGGGCTTTTGTGGAAGAGATG 460

Db 420 catccagggg 429

Qy 461 CATCCAGGGG 470

RESULT 8  
LOCUS AA111285 476 bp mRNA EST 06-NOV-1996  
DEFINITION mos3h02.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus  
CDNA clone 557331 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING  
ATPASE.1.  
ACCESSION AA111285  
NID g1663163  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Mus.

REFERENCE 1 (bases 1 to 476)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:338123  
Seq primer: -28M13 rev1 from Amerham  
High quality sequence stop: 455.  
FEATURES  
Location/Qualifiers  
1..476  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/note="Vector: pCMV-SPORT2; Site\_1: SalI; Site\_2: NotI;  
Cloned unidirectionally. Primer: Oligo dt. 10.5dpc  
embryos. pCMV-SPORT2 vector."  
/clone="557331"  
/clone\_lib="Life Tech mouse embryo 10 5dpc 10665016"  
/dev\_stage="10.5dpc embryos"  
/lab\_host="DH10B"  
BASE COUNT 92 a 149 c 147 g 88 t  
ORIGIN

Query Match 44.9%; Score 352; DB 3; Length 476;  
Best Local Similarity 89.2%; Pred. No. 0.00e+00;  
Matches 428; Conservative 0; Mismatches 47; Indels 5; Gaps 5;

Db 1 agagaaagcgaacttccgagactcctcctagcggcgcccaacctttgctccaagat 60

Qy 39 AGAGAAAGTGGCGATTCGGATCCCTAGCGCGGCCCAACCTTTACTCCAGAGAT 98

Db 61 catggtcgcgagatgtgtgctgactgcgc-gaccgagcgagtagagggcgcg 119

Qy 99 CATGGCTGCCGAGATGTGGTGGCGACTGGCGCGCACCAAGCGATCTGGAGAGCGCG 158

Db 120 gctgctgcagagatttcacgtctcctcctcaacctgctcctcctggcctgcacatt 179

Qy 159 GCTGCTGCATGAGATTTTACGTCGGCGCTCAACCTGCTGCTGCTGGCTCTGCATCTT 218

Db 180 cctgctctcaagatgcttcgagggggacacagcccggtgccagtgcgacacgacgac 239

Qy 219 CTTGCTCTCAAGATCTGCGCGGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 278

Db 240 gaa-ccaccccgctcccgccctcaagcg-cgcgacttcaacccctgcccagctgagggcg 297

Qy 279 GANGCGCGCCCTCTGCGCGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 338

Db 298 ttctgatgggtccaggaccccgcgcatctctcatggccatcaacggcaagggtgttcgacgt 357

Qy 339 CTTGCGAGCGGCTCCAGGACCGCGGCATCTCATGGCCATCAACGGCAAGGTGTCGATGT 398

Db 358 gacaaaagcgcgaagtctacgggctgaggggca-tatgggggttttcccggaagaga 416

Qy 399 GACCAAAGCGCGCAATTTACGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 458

Db 417 tgcatccagggcgcttgc-acatttgcctggacaagaagcactgaagatgagtga 475

Qy 459 TGCATCCAGGGGCTTGCCACATTTTGGCTGATAGGAAGACACTGAAGATGATGACGA 518

RESULT 9  
LOCUS AA115422 406 bp mRNA EST 15-NOV-1996  
DEFINITION z185e09.r1 Stratagene colon (#937204) Homo sapiens CDNA clone  
511432 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING ATPASE.1.  
ACCESSION AA115422  
NID g1670604  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 406)

**AUTHORS** Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.

**TITLE** WashU-Merck EST Project

**JOURNAL** Unpublished (1995)

**COMMENT** Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 326.  
Location/Qualifiers  
1..406  
/organism="Homo sapiens"  
/note="Organ: colon; Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"/>  
/clone="511432"  
/clone\_lib="Stratagene colon (#937204)"  
/lab\_host="SOLR cells (kanamycin resistant)"  
107 a 87 c 112 g 95 t 5 others

**BASE COUNT**  
mRNA  
ORIGIN  
Query Match 42.18; Score 330; DB 4; Length 406;  
Best Local Similarity 95.08; Pred. No. 0.00e+00;  
Matches 384; Conservative 0; Mismatches 12; Indels 8; Gaps 8;

Db 1 gccttcgagcggtccagaccgcgcactactatcgccatcaacgagcggttcga 60  
|||||  
Qy 336 GCCTTCGAGCGGTCCAGACCGCGCGCATCTCATGCGCATCAACGCGAGGTTCGA 395  
|||||

Db 61 tgtgacaaaggccgcaaaattctacggccgagcgcggtga-atggggtctttgtggaag 119  
|||||  
Qy 396 TGTCGACAAAGCGCAAAATTCACGGCCGCGAGGCGCGATGGGTCTTTGTGGAAG 455  
|||||

Db 120 agatgatccaggggcttgc-acattttgcctggataaggaagcactgaagatgagta 178  
|||||  
Qy 456 AGATGTCATCCAGGGGCGCTTGCACATTTTGCCTGGATAAGGAAGCACTGAAGGATGAGTA 515  
|||||

Db 179 cgatgaccttctgacctactctgcccagcaggaagactctgagtgaactggagctc 238  
|||||  
Qy 516 CGATGACCTTTCGACCTCACTGTCGCCACGAGA-GACTCTGAGTACGTGGAGTCTC 574  
|||||

Db 239 agttcatttcaagtatcatcacngtgggcaaacctgctgannngntnggggagagccac 298  
|||||  
Qy 575 AGTTCATCTTCAGATATCATCAC-GTGGGCAAACTGCTGAAGGA-GGGGAGAGGCCAC 632  
|||||

Db 299 tgtgtactcatatgaggaagaacaaagatgagagtgcccggaataattgattaaagcat 358  
|||||  
Qy 633 TGGTACTCAGATGAGGAAGAACAAAGATGAGAGTTCGCGAAAATG-TTAAAGCAT 691  
|||||

Db 359 tcagtggagatatactattttgtattttttgcaaatcatttg 402  
|||||  
Qy 692 TCAGTGAAGTATATCTATNNT-GTATTTT-GCAAAATCATTTG 733  
|||||

**RESULT** 10  
**LOCUS** AA021062 438 bp mRNA EST 30-JAN-1997  
**DEFINITION** ze67a12.r1 Soares retina N2b4HR Homo sapiens cDNA clone 364030 5'  
**ACCESSION** AA021062  
**NID** 91484823  
**KEYWORDS** EST.

**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 438)

**REFERENCE**  
**AUTHORS** Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)

**TITLE** WashU-Merck EST Project

**JOURNAL** Unpublished (1995)

**COMMENT** Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1735 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 367.  
Location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."  
/clone="364030"  
/clone\_lib="Soares retina N2b4HR"  
/sex="male"  
/tissue\_type="retina"  
/dev\_stage="55 year old"  
/lab\_host="DH10B (ampicillin resistant)"  
1..438  
mRNA 78 a 134 c 138 g 78 t 10 others  
BASE COUNT  
ORIGIN  
Query Match 39.4%; Score 309; DB 21; Length 438;  
Best Local Similarity 94.8%; Pred. No. 0.00e+00;  
Matches 364; Conservative 0; Mismatches 12; Indels 8; Gaps 8;

Db 11 gagagaaagtggcgagttccggtacccctgctagcggcccaacctttactcagaga 70  
|||||

Qy 38 GAGGAGAAAGTGGCGAGTTCGGATCCCTGCTAGCGGGCCCAACCTTTACTCCAGAGA 97  
|||||

Db 71 tcattgctgcgagagatggtggcgactggcgc-gacccaagcgatctggagagcgcg 129  
|||||

Qy 98 TCATGGCTGCCGAGATGTGGTGGCGACTGGCGCGACCCCAAGCGATCTGGAGAGCGCG 157  
|||||

Db 130 ggtcgtcgtcagagattttcacgtgcgcctcaacctctgctgtgttgctctcatct 189  
|||||

Qy 158 GGCTGCTCATGAGATTTTCAGTCGCCGCTCAACCTCTGCTGTGCTCTCATCT 217  
|||||

Db 190 tctgctctacaagatcgtgcgaggggaccagcgcgcg-cagcngacnangacgaga 248  
|||||

Qy 218 TCTTGTCTTACAGATCTGCGGGGACACAGCGCGCGCGCCAGCAGCAGCAGAGA 277  
|||||

Db 249 cga-gccgccccctctgccccgcctcaagcgcngcgacttcacccccgcgcgagctgg 307  
|||||

Qy 278 CGANGCGCCCCCTCTGCCCCGCTCAAGCGCGCGGAC-TTCACCCCGCGCGAGCTCGG 336  
|||||

```

Db 308 gtttcgacggcgtcnagggaccgcccgcatactatgccaatcaacggcaagtgttgg 367
|||||
Qy 337 CGCTTCGACGGCGTCAGG-ACCCGC-GCATACTCATGGCCATCAACGGCAAGGTGTTCG 394
|||||
Db 368 atgtgacccaagggccgcaaat 391
|||||
Qy 395 ATGTGACC-AAAGGCC-GCAAAATT 416
|||||

RESULT 11
LOCUS AA232394 349 bp mRNA EST 28-FEB-1997
DEFINITION zr27f02.r1 Strata gene NT2 neuronal precursor 937230 Homo sapiens
CDNA clone 664635 5' similar to TR:E247050 E247050 CHROMOSOME XVI
READING FRAME ORF YP1170W. ;
ACCESSION AA232394
NID 91855180
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 349)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
WashU-Merck EST Project
Unpublished (1995)
TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 revl ET from Amersham
High quality sequence stop: 214.
Location/Qualifiers
1..349
/organism="Homo sapiens"
/vector="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt
Uninduced, exponentially growing neuroepithelial cells
(Ntera-2/cl.D1). Average insert size: 1.0 kb; Uni-ZAP XR
Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
/clone="664635"
/clone_lib="Strata gene NT2 neuronal precursor 937230"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
<1..>349
61 a 122 c 108 g 58 t
mRNA
BASE COUNT
ORIGIN
Query Match 37.9%; Score 297; DB 55; Length 349;
Best Local Similarity 97.4%; Pred. No. 0.00e+00;
Matches 341; Conservative 0; Mismatches 2; Indels 7; Gaps 7;
Db 5 ccaacctttactccagagatcgtgctccgagagatgtgtgctgactgagcgc-gaccca 63
Qy 79 CCAACCTTTACTCCAGAGATCATGGCTCCGAGAGATGTGTGTGCGGACTGCGCGGCCA 138
|||||
Db 64 agcgtatcggagagcgcggtgctgcatgagattttcacgtcgcgctcaacctgctg 123
|||||
Qy 139 AGCGATCTGGAGAGCGCGGCTGCTGCATGAGATTTTTCACGTGCGGCTCAACCTGCTG 198
|||||
Db 124 ctgcttgcccttgcattctctctacaagatcgtgcgggggaccagcgcgccg 182
|||||

```

```

Qy 199 CTGCTTGGCCTCTGCAATCTTCTCTACAGATCGTGGCGGGGACACGCGGGGCC 258
Db 183 agcgcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
|||||
Qy 259 AGCGGCGACAG-GACGACGACGANGCGCGCCCTCTGCCCCGCTCAAGGCGCGAC-T 316
|||||
Db 241 tcaccccccgcgagctgcgg-gcttcgacggcgtcagagaccgcccgcatactatggcca 299
|||||
Qy 317 TCACCCCGCGCGAGCTGCGCGCTTCGAGCGCTTCAGGACCGCGCATACTACTATGSCCA 376
|||||
Db 300 tcacgcgaaggtgttcgatgtgacccaagggccgcaaatctacggggccc 349
|||||
Qy 377 TCACCGCAGGAGTTCGATGATGACCAAGGCGCGCAATTTCTACGGGCC 426
|||||

RESULT 12
ID HS1150107 standard; RNA; EST; 349 BP.
AC AA232394;
NI 91855180
DT 05-MAR-1997 (Rel. 51, Created)
DE zr27f02.r1 Strata gene NT2 neuronal precursor 937230 Homo sapiens
DE CDNA clone 664635 5' similar to TR:E247050 E247050 CHROMOSOME XVI
DE READING FRAME ORF YP1170W. ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP 1-349
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevaskis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RT "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@wustl.edu This clone is available royalty-free through
CC LNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Possible reversed clone: similarity on wrong
CC strand Seq primer: -28ml3 revl ET from Amersham High quality
CC sequence stop: 214.
FH Key Location/Qualifiers
FT source 1..349
FT /organism="Homo sapiens"
FT /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
FT XhoI; Cloned unidirectionally. Primer: Oligo dt. Uninduced
FT exponentially growing neuroepithelial cells
FT (Ntera-2/cl.D1). Average insert size: 1.0 kb; Uni-ZAP XR
FT Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
FT adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
FT /clone="664635"
FT /clone_lib="Strata gene NT2 neuronal precursor 937230"
FT /dev_stage="Ntera-2 neuroepithelial cells"
FT /lab_host="SOLR (kanamycin resistant)"
FT <1..>349
FT mRNA
SQ Sequence 349 BP; 61 A; 122 C; 108 G; 58 T; 0 other;
Query Match 37.9%; Score 297; DB 82; Length 349;
Best Local Similarity 97.4%; Pred. No. 0.00e+00;
Matches 341; Conservative 0; Mismatches 2; Indels 7; Gaps 7;
Db 5 ccaacctttactccagagatcgtgctccgagagatgtgtgctgactgagcgc-gaccca 63
Qy 79 CCAACCTTTACTCCAGAGATCATGGCTCCGAGAGATGTGTGTGCGGACTGCGCGGCCA 138
|||||
Db 64 agcgtatcggagagcgcggtgctgcatgagattttcacgtcgcgctcaacctgctg 123
|||||
Qy 139 AGCGATCTGGAGAGCGCGGCTGCTGCATGAGATTTTTCACGTGCGGCTCAACCTGCTG 198
|||||

```





```

QY 199 CTGCTGGCCTCTGCATCTTCTCTCTACAGATCGTGCAGGGGACCAAGCGCGGCC 258
Db 204 agcgat-ac--gncgacgacga-gcgccccccgctgcccccttaagcgcgcgacttc 259
QY 259 ACGGGACAGGACGACGACGAGCGCGCCCTCTGCCCCGCTCAAGCGCGCGGACTTC 318
Db 260 accctgcgagctgcgtgccttcgaagggtcccaaggaccgcgcgtatactactatgcca 319
QY 319 ACCCGCGCGAGCTGGCGGCTT-CGACGG-CGTCCAGGACCGCGGCATCTCATGSCCA 376
Db 320 tcaacggcgaaggttcgcnacgtgancgaaggccgaagttctaaagggcccaaggggc 379
QY 377 TCAACGGCAGGTTTCG-ATGTGACCAAGGCGCAATTCAT-CGGGCCCA-GGGC 433
Db 380 cgtacgggggtcttctggaagagcgcacccaagg 417
QY 434 CGTAT-GGGGTCTTCTGGAAGAGATGATCCAGGG 470

RESULT 15
LOCUS AA088127 381 bp mRNA EST 15-FEB-1997
DEFINITION mn9a10.r1 Stragene mouse Tcell 937311 Mus musculus cDNA clone
551226 5', similar to TR:E247050 E247050 CHROMOSOME XVI READING
FRAME ORF YPL170W. ;.
ACCESSION AA088127
NID g1629719
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 381)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:332018
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 336.
FEATURES
Location/Qualifiers
source
1..381
/organism="Mus musculus"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. M30 CD4+
cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
/clone="551226"
/clone_lib="Stratagene mouse Tcell 937311"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
<1..>381
mRNA 65 a 131 c 115 g 70 t
BASE COUNT
ORIGIN
Query Match 35.6%; Score 279; DB 55; Length 381;
Best Local Similarity 89.2%; Pred. No. 0.00e+00;
Matches 339; Conservative 0; Mismatches 37; Indels 4; Gaps 4;

```

```

Db 5 gtccggagctctgcttagcgggcccacactttgtctccagagatcatggtgccgagga 64
QY 54 GTTCCGGATCCCTGCCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGCTGCCGAGGA 113
Db 65 tgggtggcactgggcg-cacccgagcagcagtagaaggcggcggtgtgtgcacgagat 123
QY 114 TGTGTTGGCGACTGGCGCGCCAGCCAGCGATCTGGAGAGCGCGGGTGTGTCATGAGAT 173
Db 124 ttcaagctctccctcaacctgctctctgggcctctgcatctctctctctacaagat 183
QY 174 TTTCACGTCCCGCTCAACCTGCTGTGCTTGGCCTCTGCATCTTCTGCTCTACAAGAT 233
Db 184 cgttcgggggacacgccccggtgcccagtcgacaaacgacgacgacgaa-caccccccgct 242
QY 234 CGTGGCGGGGACACAGCGCGCGCCAGCGGACAGGACGACGACGANGCGGCCCTCT 293
Db 243 gcccgccctcaagcg-cgcgacttccacctccacctgccagctgagggcgttctgagtgctcc 301
QY 294 GCCCGGCTCAAGCGCGCGGACTT-CACCCCGCGCGAGCTGCGGCGCTTCGACGGCGTCC 352
Db 302 agaccgcgcattctcatgcccataacgcaaggtgttcgagctgacacaaaggccgca 361
QY 353 AGGACCGCGCATCTCATGCGCATCAACGCGAAGGTTCGATGTGACCAAGGCCGCA 412
Db 362 agttctacggcgctgagggg 381
QY 413 AATTCTACGGGCGGAGGGG 432

```

Search completed: Fri Nov 7 15:31:23 1997  
Job time : 995 secs.

```

OMPArch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 09:50:30 1997; Maspar time 3.07 seconds
350,234 Million cell updates/sec
Regular output not generated.

```

```
>US-08-822-364-1
Title:
Description: (1-220) from US0882264. pep
Perfect Score: 1541
Sequence: 1 MAADVATGADPDSLEGG.....SISXIFAKSFVTHVXVF 220
```

```

Scoring table:  PAM 150
                  Gap 11

Searched:      55063 seqs, 4883187 residues

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

```

```

Database:
a-issued
1:back1 2:51 3:52 4:53 5:54 6:55 7:56 8:PCr90 9:PCr91
10:PCr92 11:PCr93 12:PCr94 13:PCr95 14:PCr96

Statistics:
Mean 29.946; Variance 129.260; scale 0.232

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |         |      |        |            |    |              |            |           |
|------------|-------|---------|------|--------|------------|----|--------------|------------|-----------|
| Result No. | Score | Query % |      | Length | DB         | ID | Description  | Applicatio | Pred. No. |
|            |       | Match   |      |        |            |    |              |            |           |
| 1          | 88    | 5.7     | 311  | 5      | US-08-078- |    | Sequence 2,  | Applicatio | 9.81e+00  |
| 2          | 86    | 5.6     | 535  | 7      | US-08-208- |    | Sequence 10, | Applicati  | 1.37e+01  |
| 3          | 86    | 5.6     | 535  | 5      | US-07-906- |    | Sequence 10, | Applicati  | 1.37e+01  |
| 4          | 86    | 5.6     | 535  | 7      | US-08-167- |    | Sequence 10, | Applicati  | 1.37e+01  |
| 5          | 85    | 5.5     | 309  | 5      | US-08-078- |    | Sequence 5,  | Applicatio | 1.62e+01  |
| 6          | 85    | 5.5     | 310  | 5      | US-08-078- |    | Sequence 3,  | Applicatio | 1.62e+01  |
| 7          | 85    | 5.5     | 313  | 5      | US-08-078- |    | Sequence 4,  | Applicatio | 1.62e+01  |
| 8          | 84    | 5.5     | 485  | 12     | PCR-US94-0 |    | Sequence 2,  | Applicatio | 1.91e+01  |
| 9          | 85    | 5.5     | 487  | 7      | US-08-218- |    | Sequence 2,  | Applicatio | 1.62e+01  |
| 10         | 83    | 5.4     | 491  | 7      | US-08-206- |    | Sequence 4,  | Applicatio | 2.25e+01  |
| 11         | 80    | 5.2     | 3248 | 13     | PCR-US95-1 |    | Sequence 1,  | Applicatio | 3.68e+01  |
| 12         | 80    | 5.2     | 3248 | 6      | US-08-353  |    | Sequence 1,  | Applicatio | 3.68e+01  |
| 13         | 79    | 5.1     | 485  | 12     | PCR-US94-0 |    | Sequence 8,  | Applicatio | 4.33e+01  |
| 14         | 78    | 5.1     | 485  | 12     | PCR-US94-0 |    | Sequence 6,  | Applicatio | 5.09e+01  |
| 15         | 79    | 5.1     | 486  | 3      | US-07-737- |    | Sequence 2,  | Applicatio | 4.33e+01  |
| 16         | 79    | 5.1     | 486  | 4      | US-08-022- |    | Sequence 2,  | Applicatio | 4.33e+01  |
| 17         | 78    | 5.1     | 780  | 6      | US-08-188- |    | Sequence 50, | Applicati  | 5.09e+01  |
| 18         | 78    | 5.1     | 780  | 7      | US-08-332- |    | Sequence 44, | Applicati  | 5.09e+01  |
| 19         | 78    | 5.1     | 780  | 7      | US-08-332- |    | Sequence 50, | Applicati  | 5.09e+01  |
| 20         | 78    | 5.1     | 1229 | 4      | US-08-100- |    | Sequence 4,  | Applicatio | 5.09e+01  |
| 21         | 78    | 5.1     | 1229 | 7      | US-08-176- |    | Sequence 4,  | Applicatio | 5.09e+01  |
| 22         | 77    | 5.0     | 285  | 5      | US-08-149- |    | Sequence 24, | Applicati  | 5.97e+01  |

|   |   |  |
|---|---|--|
| CC  | MOLECULE TYPE:  | protein  |
| SQ  | SEQUENCE  | 311 AA; 32904 MW; 499722 CN;                       |
| <br>Query Match            5.7%; Score 88; DB 5; Length 311;<br>Best Local Similarity 31.1%; Pred. No. 9.81e+00;<br>Matches 14; Conservative 16; Mismatches 13; Indels 2; Gaps 2; |   |  |
| Dd  | 240 ATGASOSLLDRKEVGGVIAGLVGLIFAVCLVAFMLYRMKKKDE                       | 284<br>      : : : : :   : : : : :   : : : : :   : |
| Qy  | 8 ATGADPSDLESGLLHEIFTSPLNLLLGIC-I-FLLYKIVRGDQ                         | 50<br>    : : : : :   : : : : :   : : : : :   :    |
| <br>RESULT         2  |   |  |
| ID  | US-08-208-887A-10   | STANDARD; PRT; 535 AA.                             |
| XX  | xxxxxx  |  |
| DT  | 01-JAN-1900   |  |
| CC  | Sequence 10, Application US/08208887A.                                |  |
| CC  | Patient No. 5677421   |  |
| CC  | GENERAL INFORMATION:  |  |
| CC  | APPLICANT: Schlessinger, Joseph                                       |  |
| CC  | APPLICANT: Skolnick, Edward Y.  |  |
| CC  | APPLICANT: Margolis, Benjamin L.                                      |  |
| CC  | TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR               |  |
| CC  | TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSI |  |
| NE  | KINASES AND   |  |
| CC  | TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS                 |  |
| CC  | NUMBER OF SEQUENCES: 51   |  |
| CC  | CORRESPONDENCE ADDRESS:   |  |
| CC  | ADDRESSEE: PENNIE & EDMONDS   |  |
| CC  | STREET: 1155 Avenue of the Americas                                   |  |
| CC  | CITY: New York  |  |
| CC  | STATE: New York   |  |
| CC  | COUNTRY: 10036-2711   |  |
| CC  | ZIP: 10036-2711   |  |
| CC  | COMPUTER READABLE FORM:   |  |
| CC  | MEDIUM TYPE: Floppy disk  |  |
| CC  | COMPUTER: IBM PC compatible   |  |
| CC  | OPERATING SYSTEM: PC-DOS/MS-DOS                                       |  |
| CC  | SOFTWARE: Patentin Release #1.0, Version #1.30                        |  |
| CC  | CURRENT APPLICATION DATA:   |  |
| CC  | APPLICATION NUMBER: US/08/208.887A                                    |  |
| CC  | FILING DATE: 11-MAR-1994  |  |
| CC  | CLASSIFICATION: 435   |  |
| CC  | ATTORNEY/AGENT INFORMATION:   |  |
| CC  | NAME: Coruzzi, Laura A.   |  |
| CC  | REGISTRATION NUMBER: 30,742   |  |
| CC  | REFERENCE/DOCKET NUMBER: 7683-063                                     |  |
| CC  | TELECOMMUNICATION INFORMATION:  |  |
| CC  | TELEPHONE: (212) 790-9090   |  |
| CC  | TELEFAX: (212) 869-9741/8864  |  |
| CC  | TELEX: 66141 PENNIE   |  |
| CC  | INFORMATION FOR SEQ ID NO: 10:  |  |
| CC  | SEQUENCE CHARACTERISTICS:   |  |
| CC  | LENGTH: 535 amino acids   |  |
| CC  | TYPE: amino acid  |  |
| CC  | TOPOLOGY: unknown   |  |
| CC  | MOLECULE TYPE: protein  |  |
| SQ  | SEQUENCE  | 535 AA; 59973 MW; 1438700 CN;                      |
| <br>Query Match            5.6%; Score 86; DB 7; Length 535;<br>Best Local Similarity 36.4%; Pred. No. 1.37e+01;<br>Matches 16; Conservative 12; Mismatches 12; Indels 4; Gaps 3; |   |  |
| Dd  | 283 ESNVVVTQGRKLGMPTDFGFCVKPNKLRNGHKGHLIFCSDE                         | 326<br>:::             : : : : :   : : : : :   :   |
| Qy  | 94 NGKFVDVTKGRFYG-P-EGPYGVFAG--RDASRGLATFCLDK                         | 133<br>:::             : : : : :   : : : : :   :   |
| <br>RESULT         4  |   |  |
| ID  | US-08-167-035-10  | STANDARD; PRT; 535 AA.                             |
| XX  | xxxxxx  |  |
| DT  | 01-JAN-1900   |  |
| CC  | Sequence 10, Application US/08167035.                                 |  |
| CC  | Patient No. 5618691   |  |
| CC  | GENERAL INFORMATION:  |  |
| CC  | APPLICANT: Schlessinger, Joseph                                       |  |

|        |   |   |
|--------|---|---|
| CC     |   | COMPUTER READABLE FORM:   |
| CC     |   | MEDIUM TYPE: Floppy disk  |
| CC     |   | COMPUTER: IBM PC compatible   |
| CC     |   | OPERATING SYSTEM: PC-DOS/MS-DOS   |
| CC     |   | SOFTWARE: ASCII (text)  |
| CC     |   | CURRENT APPLICATION DATA:   |
| CC     |   | APPLICATION NUMBER: US/08/078,683A  |
| CC     |   | FILING DATE: 17-JUN-1993  |
| CC     |   | CLASSIFICATION: 435   |
| CC     |   | ATTORNEY/AGENT INFORMATION:   |
| CC     |   | NAME: Vincent, Matthew P.   |
| CC     |   | REGISTRATION NUMBER: 36,709   |
| CC     |   | REFERENCE/DOCKET NUMBER: CME-062  |
| CC     |   | TELECOMMUNICATION INFORMATION:  |
| CC     |   | TELEPHONE: (617) 227-7400   |
| CC     |   | TELEFAX: (617) 227-5941   |
| CC     |   | INFORMATION FOR SEQ ID NO: 5:   |
| CC     |   | SEQUENCE CHARACTERISTICS:   |
| CC     |   | LENGTH: 309 amino acids   |
| CC     |   | TYPE: amino acid  |
| CC     |   | TOPOLOGY: linear  |
| CC     |   | MOLECULE TYPE: peptide  |
| CC     |   | FRAGMENT TYPE: internal   |
| CC     |   | SEQUENCE 309 AA; 32678 MW; 489451 CN;   |
| SQ     |   |   |
|        |   | Query Match 5.5%; Score 85; DB 5; Length 309;<br>Best Local Similarity 28.9%; Pred. No. 1.62e+01;<br>Matches 13; Conservative 17; Mismatches 13; Indels |
| Dd     | 238   | ATGASGGLLDREVLGGVTAGLVGLIFAVCLVGFMLYRMKKKDE 282<br>     :  :  :    :  :    :  :    :  :   |
| Qy     | 8   | ATGADPSDLESGLLHEIFTPLNLLLLGLC-I-FLLYKIVRGDQ 50  |
| RESULT | 6   |   |
| ID     | US-08-078-683A-3                                      | STANDARD; PRT: 310 AA.  |
| XX     | xxxxxx  |   |
| AC     |   |   |
| DT     | 01-JAN-1900   |   |
| XX     |   |   |
| DE     | Sequence 3, Application US/08078683A.                 |   |
| XX     |   |   |
| CC     | Sequence 3, Application US/08078683A                  |   |
| CC     | Patent No. 5486599                                    |   |
| CC     | GENERAL INFORMATION:                                  |   |
| CC     | APPLICANT: Saunders, Scott                            |   |
| CC     | APPLICANT: Bernfield, Merton                          |   |
| CC     | APPLICANT: Kato, Masato                               |   |
| CC     | TITLE OF INVENTION: Construction and Use of Synthetic |   |
| CC     | TITLE OF INVENTION: Constructs Encoding Syndecan      |   |
| CC     | NUMBER OF SEQUENCES: 43                               |   |
| CC     | CORRESPONDENCE ADDRESS:                               |   |
| CC     | ADDRESSEE: LAHIVE & COCKFIELD                         |   |
| CC     | STREET: 60 State Street                               |   |
| CC     | CITY: Boston  |   |
| CC     | STATE: MA   |   |
| CC     | COUNTRY: USA  |   |
| CC     | ZIP: 02109  |   |
| CC     | COMPUTER READABLE FORM:                               |   |
| CC     | MEDIUM TYPE: Floppy disk                              |   |
| CC     | COMPUTER: IBM PC compatible                           |   |
| CC     | OPERATING SYSTEM: PC-DOS/MS-DOS                       |   |
| CC     | SOFTWARE: ASCII (text)                                |   |
| CC     | CURRENT APPLICATION DATA:                             |   |
| CC     | APPLICATION NUMBER: US/08/078,683A                    |   |
| CC     | FILING DATE: 17-JUN-1993                              |   |
| CC     | CLASSIFICATION: 435                                   |   |
| CC     | ATTORNEY/AGENT INFORMATION:                           |   |
| CC     | NAME: Vincent, Matthew P.                             |   |
| CC     | REGISTRATION NUMBER: 36,709                           |   |
| CC     | REFERENCE/DOCKET NUMBER: CME-062                      |   |
| CC     | TELECOMMUNICATION INFORMATION:                        |   |

CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 3:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 310 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 310 AA: 32492 MW: 480973 CN;

Query Match 5.5%; Score 85; DB 5; Length 310;  
 Best Local Similarity 28.9%; Pred. No. 1.62e+01;  
 Matches 13; Conservative 17; Mismatches 13; Indels 2; Gaps 2;

Db 239 ATGASQGLDRKEVLGGVLAGLVGFVAVCLVAFVFLYRMKKKDE 283  
 QY 8 ATGADPSDLESGLLHEIFTSPLNLLGLC-I-FLLYKIVRGDQ 50

RESULT 7  
 US-08-078-683A-4 STANDARD; PRT; 313 AA.  
 XX XXXXXX  
 AC XXXXXX  
 DT 01-JAN-1900  
 DE Sequence 4, Application US/08078683A.

CC Sequence 4, Application US/08078683A  
 CC Patent No. 5486599  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Saunders, Scott  
 CC APPLICANT: Bernfield, Merton  
 CC APPLICANT: Kato, Masato  
 CC TITLE OF INVENTION: Construction and Use of Synthetic  
 CC NUMBER OF SEQUENCES: 43  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & COCKFIELD  
 CC STREET: 60 State Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02109

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: ASCII (text)  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/078,683A  
 CC FILING DATE: 17-JUN-1993  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Vincent, Matthew P.  
 CC REGISTRATION NUMBER: 36,709  
 CC REFERENCE/DOCKET NUMBER: CME-062  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 4:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 313 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 313 AA: 33213 MW: 500523 CN;

Query Match 5.5%; Score 85; DB 5; Length 313;  
 Best Local Similarity 28.9%; Pred. No. 1.62e+01;  
 Matches 13; Conservative 17; Mismatches 13; Indels 2; Gaps 2;

Db 242 ATGASQGLDRKEVLGGVLAGLVGFVAVCLVAFVFLYRMKKKDE 286  
 QY 8 ATGADPSDLESGLLHEIFTSPLNLLGLC-I-FLLYKIVRGDQ 50

RESULT 8  
 ID PCT-US94-03437-2 STANDARD; PRT; 485 AA.  
 XX XXXXXX  
 AC XXXXXX  
 DT 01-JAN-1900

DE Sequence 2, Application PC/TUS9403437.  
 DE Sequence 2, Application PC/TUS9403437  
 CC GENERAL INFORMATION:  
 CC APPLICANT:

CC APPLICANT:  
 CC TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT  
 CC TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES  
 CC NUMBER OF SEQUENCES: 124  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US94/03437  
 CC FILING DATE:  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 485 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC HYPOTHETICAL: NO  
 CC ANTI-SENSE: NO  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: Leuconostoc mesenteroides  
 CC STRAIN: ATCC 12291  
 CC SEQUENCE 485 AA: 54309 MW: 1141947 CN;

Query Match 5.5%; Score 84; DB 12; Length 485;  
 Best Local Similarity 20.2%; Pred. No. 1.91e+01;  
 Matches 18; Conservative 29; Mismatches 38; Indels 4; Gaps 4;

Db 97 EAIEEAADKFDIDGNRIFVMSVAPREFGTIAKYLKSEGLADTGYNRLMIETK-PFGTISYD 155  
 QY 82 DGVQDPRIILMAING-KVFDVTKGRKEYGPGYGVFAGRDASRGLATFCIDKALKDEYD 140

Db 156 TAAELQNDLENAFDD-NQLFRID-HYLKG 182  
 QY 141 DLSDLTAAQOETLSDWESQFTFKYHHVGK 169

RESULT 9  
 ID US-08-218-943-2 STANDARD; PRT; 487 AA.  
 XX XXXXXX  
 AC XXXXXX  
 DT 01-JAN-1900

DE Sequence 2, Application US/08218943.  
 DE Sequence 2, Application US/08218943  
 CC Patent No. 5614193  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Schmaljohn, Connie S.  
 CC APPLICANT: McClain, David J.  
 CC APPLICANT: Dalrymple, Joel  
 CC APPLICANT: Dalrymple, Lonnie

CC TITLE OF INVENTION: HANTAVIRUS VACCINE  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 3000 K Street, N.W., Suite 500  
CC CITY: Washington, D.C.  
CC COUNTRY: USA  
CC ZIP: 20007-5109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/218,943  
CC FILING DATE: 28-MAR-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/799,479  
CC FILING DATE: 14-NOV-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 71007/118/USGO  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202)672-5300  
CC TELEFAX: (202)672-5399  
CC TELEX: 904136  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 487 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 487 AA; 53968 MW; 1277955 CN;

Query Match 5.5%; Score 85; DB 7; Length 487;  
Best Local Similarity 37.9%; Pred. No. 1.62e+01;  
Matches 11; Conservative 9; Mismatches 8; Indels 1; Gaps 1;

Db 446 SEWISGIFSGNW-IVLVLCVFLSLV 473  
QY 18 SGGLLHEIFTPNLNLLGLGICFLYKIV 46

RESULT 10  
ID US-08-206-176-4 STANDARD; PRT; 491 AA.  
XX  
XX  
XX 01-JAN-1900  
DE Sequence 4, Application US/08206176.  
XX  
XX Sequence 4, Application US/08206176  
XX Patent No. 5639940  
CC GENERAL INFORMATION:  
CC APPLICANT: Garner, Ian  
CC APPLICANT: Dairymple, Michael A  
CC APPLICANT: Prunkard, Donna E  
CC APPLICANT: Foster, Donald C  
CC TITLE OF INVENTION: Production of Fibrinogen in Transgenic  
CC NUMBER OF SEQUENCES: 27  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: ZymoGenetics, Inc.  
CC STREET: 4225 Roosevelt Way, N.E.  
CC CITY: Seattle  
CC STATE: WA  
CC COUNTRY: USA  
CC ZIP: 98105  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/206,176  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Parker, Gary E  
CC REGISTRATION NUMBER: 31-648  
CC REFERENCE/DOCKET NUMBER: 93-15  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 206-547-8080 ext 322  
CC TELEFAX: 206-548-2329  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 491 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 491 AA; 55928 MW; 1244445 CN;

Query Match 5.4%; Score 83; DB 7; Length 491;  
Best Local Similarity 36.7%; Pred. No. 2.25e+01;  
Matches 18; Conservative 10; Mismatches 16; Indels 5; Gaps 5;

Db 17 HLLLLLCVFLV-K-SQVNDNEEGFFSARGHRPLDK-KREE-APS-LR 60  
QY 31 NLLLLGLCIFLLYKIVRGDQPAASGDRRTTTPPLPLKRRDFTPAELR 79

RESULT 11  
ID PCT-US95-16216-1 STANDARD; PRT; 3248 AA.  
XX  
XX  
XX 01-JAN-1900  
DE Sequence 1, Application PC/TUS9516216.  
XX  
XX Sequence 1, Application PC/TUS9516216  
CC GENERAL INFORMATION:  
CC APPLICANT: Yen, Timothy J.  
CC APPLICANT: Ratner, Jerome B.  
CC TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
CC STREET: 1601 Market Street Suite 720  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: USA  
CC ZIP: 19103-2307  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US95/16216  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/353,700  
CC FILING DATE: 09-DEC-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Reed, Janet E.  
CC REGISTRATION NUMBER: 36,252  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (215) 563-4100  
CC TELEFAX: (215) 563-4044  
CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 3248 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: not relevant  
CC TOPOLOGY: not relevant  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
SQ SEQUENCE 3248 AA; 372207 MW; 51689535 CN;

Query Match 5.2%; Score 80; DB 13; Length 3248;  
Best Local Similarity 28.9%; Pred. No. 3.68e+01;  
Matches 13; Conservative 12; Mismatches 19; Indels 1; Gaps 1;

Db 2289 EMARSLKIFELDLVLRSEKENLTKIQIEKQGLSGLDCLLSFK 2333  
QY 120 DASRGLATFCLDKALKDEYDLSLTAQAQETLSDWESQFT-FK 163

RESULT 12  
ID PCT-US94-03437-1 STANDARD; PRT; 3248 AA.  
XX AC xxxxxx  
XX DT 01-JAN-1900  
XX DE Sequence 1, Application US/08353700.  
XX DE Sequence 1, Application US/08353700.  
XX DE Patent No. 5599919  
XX DE GENERAL INFORMATION:  
XX DE APPLICANT: YEN, TIMOTHY J.  
XX DE APPLICANT: RATNER, JEROME B.  
XX DE TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
XX DE TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
XX DE TITLE OF INVENTION: AND METHODS OF USE  
XX DE NUMBER OF SEQUENCES: 4  
XX DE CORRESPONDENCE ADDRESSES:  
XX DE ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
XX DE STREET: 1601 MARKET STREET, SUITE 720  
XX DE CITY: PHILADELPHIA  
XX DE STATE: PA  
XX DE COUNTRY: USA  
XX DE ZIP: 19103-2307  
XX DE COMPUTER READABLE FORM:  
XX DE MEDIUM TYPE: Floppy disk  
XX DE OPERATING SYSTEM: PC-DOS/MS-DOS  
XX DE SOFTWARE: Patent in Release #1.0, Version #1.25  
XX DE CURRENT APPLICATION DATA:  
XX DE FILING DATE: 09-DEC-1994  
XX DE CLASSIFICATION: 435  
XX DE ATTORNEY/AGENT INFORMATION:  
XX DE NAME: REED, JANET E.  
XX DE REGISTRATION NUMBER: 36,252  
XX DE TELECOMMUNICATION INFORMATION:  
XX DE TELEPHONE: (215) 563-4100  
XX DE TELEFAX: (215) 563-4044  
XX DE INFORMATION FOR SEQ ID NO: 1:  
XX DE SEQUENCE CHARACTERISTICS:  
XX DE LENGTH: 3248 amino acids  
XX DE TYPE: amino acid  
XX DE STRANDEDNESS: single  
XX DE TOPOLOGY: linear  
XX DE MOLECULE TYPE: protein  
XX DE HYPOTHETICAL: NO  
XX DE ANTI-SENSE: NO  
XX DE ORIGINAL SOURCE: HUMAN  
XX DE ORGANISM: HUMAN  
SQ SEQUENCE 3248 AA; 372207 MW; 51689535 CN;

Query Match 5.2%; Score 80; DB 6; Length 3248;

Best Local Similarity 28.9%; Pred. No. 3.68e+01;  
Matches 13; Conservative 12; Mismatches 19; Indels 1; Gaps 1;

Db 2289 EMARSLKIFELDLVLRSEKENLTKIQIEKQGLSGLDCLLSFK 2333  
QY 120 DASRGLATFCLDKALKDEYDLSLTAQAQETLSDWESQFT-FK 163

RESULT 13  
ID PCT-US94-03437-8 STANDARD; PRT; 485 AA.  
XX AC xxxxxx  
XX DT 01-JAN-1900  
XX DE Sequence 8, Application PC/TUS9403437.  
XX DE Sequence 8, Application PC/TUS9403437.  
XX DE GENERAL INFORMATION:  
XX DE APPLICANT:  
XX DE APPLICANT:  
XX DE TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT  
XX DE TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES  
XX DE NUMBER OF SEQUENCES: 124  
XX DE COMPUTER READABLE FORM:  
XX DE MEDIUM TYPE: Floppy disk  
XX DE COMPUTER: IBM PC compatible  
XX DE OPERATING SYSTEM: PC-DOS/MS-DOS  
XX DE SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
XX DE CURRENT APPLICATION DATA:  
XX DE APPLICATION NUMBER: PCT/US94/03437  
XX DE FILING DATE:  
XX DE INFORMATION FOR SEQ ID NO: 8:  
XX DE SEQUENCE CHARACTERISTICS:  
XX DE LENGTH: 485 amino acids  
XX DE TYPE: amino acid  
XX DE STRANDEDNESS: single  
XX DE TOPOLOGY: linear  
XX DE MOLECULE TYPE: protein  
XX DE HYPOTHETICAL: NO  
XX DE ANTI-SENSE: NO  
XX DE ORIGINAL SOURCE:  
XX DE ORGANISM: Leuconostoc dextranum  
XX DE STRAIN: ATCC 19255  
SQ SEQUENCE 485 AA; 54443 MW; 1139004 CN;

Query Match 5.1%; Score 79; DB 12; Length 485;  
Best Local Similarity 19.3%; Pred. No. 4.33e+01;  
Matches 17; Conservative 28; Mismatches 39; Indels 4; Gaps 4;

Db 98 ATEEAATKFDIDGNRIFYMSVAPRFGTTAKYKLSGLLAETGYNRLMIEK-PFGTSYAT 156  
QY 83 GVQDPRIILMAING-KVFDVTGKRKYGPEGYGVFAGRDASRGLATFCLDKALKDEYD 141

Db 157 ABELQSDLENAFDD-DQLFRID-HYLGR 182  
QY 142 LSDLTAQAQETLSDWESQFTFYHVGK 169

RESULT 14  
ID PCT-US94-03437-6 STANDARD; PRT; 485 AA.  
XX AC xxxxxx  
XX DT 01-JAN-1900  
XX DE Sequence 6, Application PC/TUS9403437.  
XX DE Sequence 6, Application PC/TUS9403437.  
XX DE GENERAL INFORMATION:  
XX DE APPLICANT:  
XX DE APPLICANT:  
XX DE TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT



CC TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES  
CC NUMBER OF SEQUENCES: 124  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/03437  
CC FILING DATE:  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 485 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORGANISM: Leuconostoc lactis  
CC STRAIN: NCDO 546  
CC SEQUENCE 485 AA; 54253 MW; 1136158 CN;

Query Match 5.1%; Score 78; DB 12; Length 485;  
Best Local Similarity 19.3%; Pred. No. 5.09e+01;  
Matches 17; Conservative 29; Mismatches 38; Indels 4; Gaps 4;  
Db 98 AIEEASDKFGIDGNRIFYMSVAPREFGTIAKYLKSEGLLATTGYNRLMIK-PEGTSYET 156  
QY 83 GVQDPRIILMAING-KVFDVTGKRFGYGPYGVFAGRDASRLATFCLDKREALKDEYDD 141  
Db 157 AEKLQLENAFDD-DQLFRID-HYLK 182  
QY 142 LSDLTAQQETLSDWESQFTFKYHHVCK 169

RESULT 15  
ID US-07-737-071A-2 STANDARD; PRT; 486 AA.  
XX xxxxxx  
XX  
DT 01-JAN-1900  
DE Sequence 2, Application US/07737071A.  
CC Sequence 2, Application US/07737071A  
CC Patent No. 5229286  
CC GENERAL INFORMATION:  
CC APPLICANT: JARSCH, Michael  
CC APPLICANT: LANG, Gunter  
CC TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF  
CC TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTO  
CC  
CC TITLE OF INVENTION: DEXTRANICUS  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Armstrong, Nikola, Marmelstein Kubovcik &  
CC ADDRESSEE: Murray  
CC STREET: 1725 K Street, N.W., Suite 1000  
CC CITY: Washington D.C.  
CC COUNTRY: United States of America  
CC ZIP: 20006  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/737,071A  
CC FILING DATE: 19910730  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: DE P 4024158.9

CC FILING DATE: 30-JUL-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Murray, Robert B.  
CC REGISTRATION NUMBER: 22,890  
CC REFERENCE/DOCKET NUMBER: 911028  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 659-2930  
CC TELEFAX: (202) 887-0357  
CC TELEX: 440142  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 486 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 486 AA; 54575 MW; 1143728 CN;  
Query Match 5.1%; Score 79; DB 3; Length 486;  
Best Local Similarity 19.3%; Pred. No. 4.33e+01;  
Matches 17; Conservative 28; Mismatches 39; Indels 4; Gaps 4;  
Db 99 AIEEATKFDIDGNRIFYMSVAPREFGTIAKYLKSEGLLAETGYNRLMIK-PEGTSYAT 157  
QY 83 GVQDPRIILMAING-KVFDVTGKRFGYGPYGVFAGRDASRLATFCLDKREALKDEYDD 141  
Db 158 AEELQSDLENAFDD-DQLFRID-HYLK 183  
QY 142 LSDLTAQQETLSDWESQFTFKYHHVCK 169  
Search completed: Thu Nov 6 09:50:44 1997  
Job time: 14 secs.

\*\*\*\*\*

W P S R E A  
(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Nov 7 14:43:51 1997; MasPar time 682.96 Seconds  
Tabular output not generated. 1267.185 Million cell updates/sec

Title: >US-08-822-264-2  
Description: (1-788) from US08822264.seq  
Perfect Score: 784  
N.A. Sequence: 1 GCCGCCGAACCCGCGCGCC.....AATATTAGAAAGTTTGAGC 788  
Comp: CGCGCGCTTGGCGCGCGG.....TTATAATCTTCAACTCG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-bcwt  
1:BCWT 3:GEN1 4:GEN2 5:HTG1 6:HTG2 7:HUM 8:INV  
9:ORG 10:MAM 11:VRT 12:PLN 13:PRO 14:ROD 15:SYN 16:UNC  
17:VIR

Database: genbank99  
18:BC11 19:BC12 20:BC13 21:BC14 22:BC15 23:BC16 24:BC17  
25:BC18 26:BC19 27:BC10 28:BC11 29:GEN1 30:GEN2  
31:GEN3 32:HTG1 33:HTG2 34:HTG3 35:INV1 36:INV2 37:INV3  
38:INV4 39:INV5 40:INV6 41:INV7 42:INV8 43:INV9 44:INV10  
45:INV11 46:MAM1 47:MAM2 48:MAM3 49:VRT1 50:VRT2 51:VRT3  
52:VRT4 53:PAT1 54:PAT2 55:PAT3 56:PAT4 57:PAT5 58:PHG  
59:PLN1 60:PLN2 61:PLN3 62:PLN4 63:PLN5 64:PLN6 65:PLN7  
66:PLN8 67:PLN9 68:PLN10 69:PLN11 70:PLN12  
72:PRI3 73:PRI4 74:PRI5 75:PRI6 76:PRI7 77:PRI8 78:PRI9  
79:PRI10 80:PRI11 81:PRI12 82:PRI13 83:PRI14 84:PRI15  
85:ROD1 86:ROD2 87:ROD3 88:ROD4 89:ROD5 90:ROD6 91:ROD7  
92:ROD8 93:STR 94:SYN 95:UNA 96:VRL1 97:VRL2 98:VRL3  
99:VRL4 100:VRL5 101:VRL6 102:VRL7 103:VRL8 104:VRL9  
105:VRL10  
genbank-new3  
106:BC1 107:GEN1 108:GEN2 109:HTG1 110:HTG2 111:INV  
112:NAM 113:VRT 114:PHG 115:PLN 116:PRI1 117:PRI2  
118:ROD 119:SYN 120:UNA 121:VRL  
u-emb150\_99  
122:part1

Database:

Database: u-emb150\_99  
122:part1

Statistics: Mean 11.018; Variance 4.807; scale 2.292

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description           | Pred. No. |
|------------|-------|-------------|--------|-------|-----------------------|-----------|
| 1          | 563   | 71.8        | 1893   | 48    | SSSTERMBP             | 0.00e+00  |
| 2          | 487   | 62.1        | 1885   | 92    | RND63315              | 0.00e+00  |
| 3          | 40    | 5.1         | 215    | 57    | Sequence 5 from paten | 4.18e-11  |
| c 4        | 31    | 4.0         | 215    | 57    | Sequence 5 from paten | 8.72e-05  |
| c 5        | 29    | 3.7         | 1540   | 48    | S.scrofa mRNA for BM8 | 1.75e-03  |
| 6          | 28    | 3.6         | 354    | 8     | Oxytricha fallax 57KD | 7.55e-03  |
| 7          | 28    | 3.6         | 354    | 111   | Oxytricha fallax 57KD | 7.55e-03  |
| 8          | 28    | 3.6         | 6810   | 92    | Rattus mRNA for bra   | 7.55e-03  |
| 9          | 27    | 3.4         | 201    | 53    | S.griseus gene for pr | 3.16e-02  |
| 10         | 27    | 3.4         | 201    | 53    | Synthetic DNA for pr  | 3.16e-02  |
| c 11       | 27    | 3.4         | 201    | 53    | Synthetic DNA for pr  | 3.16e-02  |
| c 12       | 27    | 3.4         | 201    | 53    | S.griseus gene for pr | 3.16e-02  |
| c 13       | 26    | 3.3         | 827    | 63    | Podospora anserina mi | 1.28e-01  |
| c 14       | 26    | 3.3         | 1739   | 63    | Podospora anserina mi | 1.28e-01  |
| c 15       | 26    | 3.3         | 4061   | 41    | D.melanogaster putat  | 1.28e-01  |
| c 16       | 26    | 3.3         | 100314 | 64    | PANMTPACGA            | 1.28e-01  |
| c 17       | 26    | 3.3         | 100314 | 63    | MPACG                 | 1.28e-01  |
| c 18       | 25    | 3.2         | 201    | 53    | S.griseus gene for pr | 5.02e-01  |
| c 19       | 25    | 3.2         | 201    | 53    | Synthetic DNA for pr  | 5.02e-01  |
| c 20       | 25    | 3.2         | 201    | 53    | Synthetic DNA for pr  | 5.02e-01  |
| c 21       | 25    | 3.2         | 201    | 53    | S.griseus gene for pr | 5.02e-01  |
| c 22       | 25    | 3.2         | 2209   | 67    | SCYPL170W             | 5.02e-01  |
| c 23       | 25    | 3.2         | 2779   | 25    | MXCARD                | 5.02e-01  |
| c 24       | 25    | 3.2         | 48096  | 38    | CET20G5               | 5.02e-01  |
| c 25       | 25    | 3.2         | 55786  | 66    | SLACHXVI              | 5.02e-01  |
| c 26       | 24    | 3.1         | 838    | 80    | HUNHAETR              | 1.90e+00  |
| c 27       | 24    | 3.1         | 1067   | 19    | CCRFLAD               | 1.90e+00  |
| c 28       | 24    | 3.1         | 1108   | 19    | CCRFINA               | 1.90e+00  |
| c 29       | 24    | 3.1         | 1557   | 47    | DOGIDUA03             | 1.90e+00  |
| c 30       | 24    | 3.1         | 2052   | 27    | STMTETA               | 1.90e+00  |
| c 31       | 24    | 3.1         | 2423   | 50    | GGU17277              | 1.90e+00  |
| c 32       | 24    | 3.1         | 2520   | 78    | HSU50383              | 1.90e+00  |
| c 33       | 24    | 3.1         | 2538   | 96    | ACZTKSEA              | 1.90e+00  |
| c 34       | 24    | 3.1         | 2982   | 18    | ATLEUCD               | 1.90e+00  |
| c 35       | 24    | 3.1         | 6716   | 26    | PSETODC1C             | 1.90e+00  |
| c 36       | 24    | 3.1         | 22069  | 46    | BTAS1C                | 1.90e+00  |
| c 37       | 24    | 3.1         | 45679  | 35    | CEC08F11              | 1.90e+00  |
| c 38       | 23    | 2.9         | 1371   | 20    | D50588                | 1.90e+00  |
| c 39       | 23    | 2.9         | 3115   | 38    | CEU34812              | 1.90e+00  |
| c 40       | 23    | 2.9         | 5000   | 85    | CRUGAPB               | 1.90e+00  |
| c 41       | 23    | 2.9         | 9134   | 77    | HSU203H4              | 1.90e+00  |
| c 42       | 23    | 2.9         | 12177  | 78    | HSU56438              | 1.90e+00  |
| c 43       | 23    | 2.9         | 13356  | 104   | RCU50550              | 1.90e+00  |
| c 44       | 23    | 2.9         | 31433  | 37    | CELF54C1              | 1.90e+00  |
| c 45       | 23    | 2.9         | 37331  | 34    | HSCB33B10             | 1.90e+00  |

ALIGNMENTS

RESULT 1 SSSTERMBP 1893 bp RNA MAM 01-NOV-1996  
LOCUS S.scrofa mRNA for steroid membrane binding protein.  
DEFINITION X99714  
ACCESSION g1657408  
NID steroid membrane binding protein.  
KEYWORDS pig.  
SOURCE Sus scrofa  
ORGANISM Eukaryotes; mitochondria; Metazoa; Chordata;  
Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 1893)  
AUTHORS Falkenstein, E.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1893)  
AUTHORS Falkenstein, E.  
TITLE Direct Submission  
JOURNAL Submitted (01-AUG-1996) E. Falkenstein, University of Heidelberg,  
Institute of Clinical Pharmacology, Faculty of Clinical Medicine at  
Mannheim, Theodor-Kutzer-Ufer, 68135 Mannheim, FRG  
COMMENT Reference: Meyer, C.; J.Biol.Chem. 239; 726-731; 1996 (N-terminal  
protein-sequence).

```
FEATURES
  source      Location/Qualifiers
  1..1893
    /organism="Sus scrofa"
    /sub_species="domestica"
    /tissue_type="aorta"
    /cell_type="smooth muscle cells"
    /clone_lib="Uni-ZAP XR-Vector (Fa. Stratagene)"
    /clone="PgMEF6.1"
    42..626
  CDS
    /codon_start=1
    /product="steroid membrane binding protein"
    /db_xref="PID:e257707"
    /db_xref="PID:g1657409"
    /translation="MAEDVATGADPSELEGGLHEFTSPINLLGLCFLLYK
    IVRGQPAASDSDPPPLRLKRFDPFAELRRFDGVQDPRILMAINGKVFDTVK
    RFYGPGEYGVYFAGDASRGATFCLDKALKDEYDLDLTPAQETLNDWDSQFT
    FXHHVKKLLKGEETVYSDDEEPKDESARKND"
  BASE COUNT  509 a  447 c  461 g  476 t
  ORIGIN
    Query Match      71.8%; Score 563; DB 48; Length 1893;
    Best Local Similarity 91.9%; Pred. No. 0.00e+00;
    Matches 647; Conservative 0; Mismatches 51; Indels 6; Gaps 5;
  Db 21 ccaaccttgcctcagcagatcggtgcgagagatgtgcggtaccgagcgccgacccg 80
  QY 79 CCAACCTTTACTCCAGAGCATGGTCCGCGAGGATGGTGGCGACTGGCGCGACCCA 138
  Db 81 agcgaactagagcgcggtgcgtgcagatgattttcagtcgcgcctcaactcgtg 140
  QY 139 AGCGATCTGGAGAGCGCGGGCTGCTGCATGAGATTTTCACGTCGCGCTCAACCTGCTG 198
  Db 141 ctgctcgccctgcctcttcctgctctcaagatcgctgcggggaccagcgcgcgcc 200
  QY 199 CTGCTGGCTCTGCATCTCTCTGCTCTACAGATCGTGGCGGGACCGCGCGGCC 258
  Db 201 agcgatagc--gacgacgacga-gcgccccctgcctcccgcttaagcgcgcgacttc 257
  QY 259 AGCGCGGACAGGACGACGAGCGCGCCCTCTGCGCGCTCAAGCGCGCGGACTTC 318
  Db 258 accctcgagctgctgctgcgagcggtccaggaccggtatactcattcagtcgcatc 317
  QY 319 ACCCCCGCGAGTGGCGGCTTCGACGGCTCCAGAGCGCTCCAGACCGCGCATCTATGGCCATC 378
  Db 318 aacggcaaggtgttcgactgacgacgaagcgcgcaagttctacggcgccgagggcgctac 377
  QY 379 AACGGCAAGGTGTTCTGATGTGACCAAGCGCGCAATTTCTACGGGCGCGAGGGCCGTAT 438
  Db 378 ggggttcttgcctggaagacgcacatccaggggcgctggccacgttttgcctggataagaa 437
  QY 439 GGGGTCTTTGCTGGAAGAGATGATCATCAGGGGCTTGCCACATTTTGCTGGATAAGGAA 498
  Db 438 gcccgaagagagatgatgacatttctgacctcaactcactcctcccaagcagagaccctg 497
  QY 499 GCACGTGAAGGATGAGTACCATGACCTTTCTGACCTCTCACCTGCTCCCGACGAGACACTG 558
  Db 498 aatgactggactctcagttcactttcaagtaaccatcagctggcgaactgctgaagag 557
  QY 559 AGTACTGGAGTCTCAGTTCACATTTCAAGTATCATCATGCTGGCAACCTGCTGAAGGAG 618
  Db 558 ggggaggagccaccctgtactcagatgaggaagagcccaagatgagagcgctcggaaa 617
  QY 619 GGGGAGGAGCCCATCTGTACTCAGATGAGGAAGAACCAAAAGATGAGATTTCCCGGAAA 678
  Db 618 aatgattaaagcgttcggtggaagcatatctattttgtattttgcagaatcatttgtaa 677
  QY 679 AATG-TTAAAGCATTCAGTGGAAAGTATCTATNT-GTATTTTGGAAAATCATTTGTAA 736
  Db 678 cattccagctcgtctctaaacatgatttcaattattagaa 721
  QY 737 CAGTCCACATGTCCTTTAAACATAGTG-TTACAATATTATTAGAA 779
```

```
RESULT
LOCUS      RNU63315      1885 bp      mRNA      ROD      02-SEP-1996
DEFINITION Rattus norvegicus 25-Dx (25Dx) mRNA, complete cds.
ACCESSION  U63315
NID        g1518817
KEYWORDS   Norway rat.
SOURCE     Rattus norvegicus
ORGANISM   Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE  1 (bases 1 to 1885)
AUTHORS   Selmin,O., Lucier,G., Clark,G., Tritscher,A., Vanden-Heuvel,J.,
            Gastel,J., Walker,N., Sutter,T. and Bell,D.A.
TITLE     Isolation and characterization of a novel gene induced by 2,3,7,8
            TCDD in rat liver
JOURNAL   Carcinogenesis (1996) In press
REFERENCE  2 (bases 1 to 1885)
AUTHORS   Bell,D.A. and Selmin,O.
TITLE     Direct Submission
JOURNAL   Submitted (09-JUL-1996) LCBRA, NIEHS, P.O. Box 12233, Research
            Triangle Park, NC 27709, USA
FEATURES   Location/Qualifiers
            1..1885
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /tissue_type="liver"
            98..769
            /gene="25DX"
            /note="25 Kd protein; gene is TCDD inducible similar to
            cytokine/growth factor/prolactin receptor superfamily"
            /codon_start=1
            /product="25-Dx"
            /db_xref="PID:g1518818"
            /translation="MAEDVATGADPSELEGGLHEFTSPINLLGLCFLLYK
            IVRGQPAASDNDPPPLRLKRFDPFAELRRFDGVQDPRILMAINGKVFDTVK
            KRFYGPGEYGVYFAGDASRGATFCLDKALKDEYDLDLTPAQETLNDWDSQF
            SSPSTTTWGLKEAEPIVSDDEEQMRLLGRVTEAVSGAYLFLYFAKSEVTFQS
            VFTTW"
BASE COUNT  550 a  399 c  427 g  509 t
ORIGIN
    Query Match      62.1%; Score 487; DB 92; Length 1885;
    Best Local Similarity 86.4%; Pred. No. 0.00e+00;
    Matches 658; Conservative 0; Mismatches 90; Indels 14; Gaps 10;
  Db 26 tcgcttagagggcgaggaagcgcgactgtttcgg-tctctgataaacagcccaacctt 84
  QY 27 TCGCTCAGAGGAGGAGAGAAAGTGGCGAGTTCGGGATCCCTGCTAGCGCGCGCAACCTT 86
  Db 85 tgcctcagagatcatgctgcgagagatgtggtggcagctggcgccagcccgagcgt 144
  QY 87 TACTCCAGAGATCATGGCTGCCGAGGATGTTGGCGGACTGGCGCCGACCAACGATCT 146
  Db 145 gggaggcgggcggtgcttcaagagattttcacgtgcctctcaacctgctccttgg 204
  QY 147 GGAGAGCGCGGGCTGCTGCATGAGATTTTCACGTCGCCCTCAACCTGCTGCTGTG 206
  Db 205 cctctgattcttctgctctacaagatcgcttcgcgggagaccgcccgggtgccagtggggga 264
  QY 207 CCTCTGATCTTCTGCTCTACAAGATCGTGGCGGGGACAGCGCGCGCGCGAGCGCGA 266
  Db 265 caacgacgacgacga-gccgcccccgctgcctcagcgcgctgaattcacccctg 323
  QY 267 CAG-GAGCAGCAGCAGCAGCCCGCCCTCTGCCCGCCCTCAAGCGCGCGGACTTCACCCCGC 325
  Db 324 ccgaactaaaggcatcacatggatggatccagaccgcccgcattcttatggccatcaacgca 383
  QY 326 CCGAGCTGCGGCGCTTCGACGGGCTTCAGGACCCGCGCATACTCATGGCCATCAACGCCA 385
  Db 384 aggtgttcgactgacccaagggcgcaagtcttatggccggaggggaccataagggtct 443
  QY 386 AGGTGTTTCGATGTGACCAAGGCGCGCAAAATCTACGGCGCGGAGGGCGCGTATGGGTCT 445
```

Db 444 ttgctgaagagatgcattccaggggcttccacatttccctggacaaagaagcactga 503  
 |||||  
 QY 446 TTGCTGGAAGAGATGCATCCAGGGGCTTCCACATTTTCCCTGGATAAGGAAGCACTGA 505  
 |||||  
 Db 504 aggatgatgatgatgacctttctgacctctactcctccagcaggagacactgaatgact 563  
 |||||  
 QY 506 AGGATGAGTACGATGACCTTTCTGACCTCTACTGCTGCCACGAGGAGACTCTGAGTGACT 565  
 |||||  
 Db 564 gggactcctcagttcagttcaccctcagttcaccatcagctgggaaactcttgaagag 623  
 |||||  
 QY 566 GGGAGTCTCAGTTCATTT-----TCAAGTATCATCATCTGCTGGGCAA-CTGCTGAAGAGG 619  
 |||||  
 Db 624 cggaggagcggattgttactcggatgatgaagaac-aaagatgagctgctcggaaga 682  
 |||||  
 QY 620 GGGAGGAGCCACTGTGTACTCAGATGAGGAAGAACCAAAAGATGAGAGTCCCGGAAAA 679  
 |||||  
 Db 683 gtgactgaagcagctagtgga-gcatatctattttgtatttgcataaatttgaac 741  
 |||||  
 QY 680 ATGT-TAAAGCATTCAGTGGAGATATATCTATNTT-GTATTTGCAAAATCATTTGTAAC 737  
 |||||  
 Db 742 attccagctgtctttacacatggtgatttcaatttaga 783  
 |||||  
 QY 738 AGTCCACTGTGCTTTAAACATAGTG-TTACATATTTAGA 778  
 |||||

RESULT 3 128278 215 bp DNA PAT 30-OCT-1996  
 LOCUS  
 DEFINITION Sequence 5 from patent US 5569830.  
 ACCESSION 128278  
 NID 91819054  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 215)  
 AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.  
 TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
 JOURNAL Patent: US 5569830-A 5 29-OCT-1996;

FEATURES source Location/Qualifiers  
 1..215  
 BASE COUNT 15 a 8 c 25 g 26 t 141 others  
 ORIGIN

Query Match 5.1%; Score 40; DB 57; Length 215;  
 Best Local Similarity 14.2%; Pred. No. 4.18e-11;  
 Matches 23; Conservative 73; Mismatches 64; Indels 2; Gaps 2;

Db 6 masssvvtasendkaktgnttsvttddcnrtgvcddtttyrnnndsgnhyssan 65  
 QY 13 CGCGCCCACTCCTCGCTCAGAGGGAGAGAAAGTCCGGATTCCTCGCTAG 72  
 |||||

Db 66 ynggnnvgaakthyythtvsagadsktvdtsynasgttdnrgsadsgysktt 125  
 |||||

QY 73 CGCGGCC-AACCTTTACTCCAGATCATGTGCTCGGAGAGATGTGTGCGGACTGCGC 131  
 |||||

Db 126 antsrnrtgtannavd-srnmmdasvsgdkntkhhaknsad 166  
 QY 132 CGACCCAAAGCATCTGGAGAGCGGGCTGCTGCATGAGAT 173  
 |||||

RESULT 4 128278 215 bp DNA PAT 30-OCT-1996  
 LOCUS  
 DEFINITION Sequence 5 from patent US 5569830.  
 ACCESSION 128278  
 NID 91819054  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 215)  
 AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.  
 TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
 JOURNAL Patent: US 5569830-A 5 29-OCT-1996;

FEATURES source Location/Qualifiers  
 1..215  
 BASE COUNT 15 a 8 c 25 g 26 t 141 others  
 ORIGIN

Query Match 4.0%; Score 31; DB 57; Length 215;  
 Best Local Similarity 14.2%; Pred. No. 8.72e-05;  
 Matches 25; Conservative 71; Mismatches 78; Indels 2; Gaps 2;

Db 40 twgvcdtdtyrvndsgnhyssanyngnnvgaakthyythtvsagadsktvdtsyn 99  
 |||||

Cp 413 TTGGGGCTTTGGTCACATCGAACACCTTGCCGTGTGATGGCCATGAGTATCGCGGGTCC 354  
 |||||

Db 100 asgtssnggtdnrgsadsgyskttamtsnrnrtgtannavdsrnmmdasvsgdkntk 159  
 |||||

Cp 353 TGGAGCGCGTCGAAGCGCGGAGCTCGCGGGG-GTGAAGTCGCGCCCTTGAGCGGGG 295  
 |||||

Db 160 haknsadgkvsknngdrnrrygtgtksnvann-cgggnkrdsasyannkccgasc 214  
 |||||

Cp 294 CAGAGGGGGCGCTCGTCGTCCTGTCGCCGCTGCGCGCGCTGGTCCCGC 239  
 |||||

RESULT 5 SSBM88 1540 bp RNA MAM 28-JUN-1995  
 LOCUS  
 DEFINITION S.scrofa mRNA for BM88 antigen.  
 ACCESSION X82027  
 NID g557672  
 KEYWORDS BM88 antigen.  
 SOURCE pig.  
 ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 1540)  
 AUTHORS Mamalaki,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-OCT-1994) A. Mamalaki, Hellenic Pasteur Institute,  
 Dept of Biochemistry, 127 Vas. Sofias Ave., 11521 Athens, GREECE

REFERENCE 2 (bases 1 to 1540)  
 AUTHORS Mamalaki,A., Boutou,E., Hurel,C., Patsavoudi,E., Tzartos,S. and Matsas,R.  
 TITLE The BM88 antigen, a novel neuron-specific molecule, enhances the differentiation of mouse neuroblastoma cells  
 JOURNAL J. Biol. Chem. 270 (23), 14201-14208 (1995)  
 MEDLINE 95294030  
 FEATURES source Location/Qualifiers  
 1..1540  
 /organism="Sus scrofa"  
 /dev\_stage="adult"  
 /tissue\_type="pig brain"  
 /clone\_lib="pig brain kgt11"  
 1..1540  
 /gene="BM88"  
 /codon\_start=1  
 /product="BM88 antigen"  
 /db\_xref="PID:g557673"

mRNA  
 CDS  
 124..546  
 /gene="BM88"

BASE COUNT 274 a 525 c 491 g 250 t  
 ORIGIN

Query Match 3.7%; Score 29; DB 48; Length 1540;  
 Best Local Similarity 74.6%; Pred. No. 1.75e-03;  
 Matches 44; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QEPAAATLPKATKALDNHSLPAPAPSPDPAEPKFGPDGAEAEAPSGGP  
 GGRGCPFFENLPLVAGGIAMAAAILGVFLARKK"

Db 1209 tctgcaattgccctggggcagctgctgctgcccagccccagagggccctcccgcc 1267  
 Cp 480 TGTGGCAAGCCCTGGATCTCTTCACGACAAAGACCCATACGGCCCTCGGCC 422

RESULT 6  
 ID OFU89259 standard; DNA; INV; 354 BP.  
 AC U89259;  
 NI 91881675  
 DT 13-MAR-1997 (Rel. 51, Created)  
 DT 13-MAR-1997 (Rel. 51, Last updated, Version 1)  
 DE Oxytricha fallax 57kd zinc finger/protein chimera gene, partial  
 DE cds.  
 KW Oxytricha fallax  
 OS Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;  
 OC hypotrichs; Stichotrichida; Oxytrichidae; Oxytricha.  
 RN [1]  
 RP 1-354  
 RP MEDLINE; 94134747.  
 RA Doak T.G., Doerder F.P., Jahn C.L., Herrick G.;  
 RA "A proposed superfamily of transposase genes: transposon-like  
 RT elements in ciliated protozoa and a common 'D35E' motif";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:942-946(1994).  
 RP [2]  
 RP 1-354  
 RA Witherspoon D.J., Doak T.G., Williams K., Seger J., Herrick G.;  
 RA "Selection on the protein-coding genes of the TBE1 family of  
 RT transposable elements in the ciliates Oxytricha fallax and O.  
 RT trifallax";  
 RL Unpublished.  
 RL [3]  
 RP 1-354  
 RA Doak T.G., Williams K., Witherspoon D.J., Herrick G.;  
 RT Submitted (11-FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RL Oncological Science, University of Utah, School of Med. Rm5C334,  
 RL USA, UT 84132, USA  
 FH Key Location/Qualifiers  
 FT source  
 FT 1..354  
 FT /organism="Oxytricha fallax"  
 FT /strain="9D1"  
 FT /transposon="TBE1"  
 FT /note="this is a bulk sequence that was generated from a  
 FT PCR product that represents many transposon templates"  
 FT <1..>354  
 FT /codon\_start=1  
 FT /product="57kd zinc finger/protein chimera"  
 FT /transl\_table=6  
 FT /db\_xref="PID:gl881676"  
 FT /translation="HTRDLXKHLKRAHKKXXEXXXXXXKXLRKAREXXXXXX  
 FT AXEXXXVXXRNXLSEXTKIMIKIQYKIPVLAQIDLTSLQSYLLEDSFDPKVI  
 FT HQYAKVYFPNYLX"  
 FT Q  
 FT D  
 FT SQ Sequence 354 BP; 106 A; 42 C; 41 G; 54 T; 111 others;  
 Query Match 3.6%; Score 28; DB 8; Length 354;  
 Best Local Similarity 26.6%; Pred. No. 7.55e-03;  
 Matches 37; Conservative 50; Mismatches 50; Indels 2; Gaps 2;  
 Db 30 aaaaagcayaraagadtagarrttgarrakhawgmhwsarytraagytdctcaa 89  
 QY 650 AAGAACCAAAAGATGAGAGTCCCGGAAAATGTTAAAGCATTTCAGTGAAGTATATCTA 709  
 Db 90 haaagaagaar-gcdagrgaryhyhdwargwb-gydgwgcyykaayaagcwagmgarwya 147  
 QY 710 TNNGTATTTTGCAAAATCATTTGTGAACAGTCCACATNTGCTTTTAAACATAGTGTACA 769  
 Db 148 swrtrygtaargawdkrc 166  
 QY 770 ATATTTAGAAAGTTTGAGC 788

QY 770 ATATTTAGAAAGTTTGAGC 788

RESULT 7  
 LOCUS  
 DEFINITION Oxytricha fallax 57kd zinc finger/protein chimera gene, partial  
 cds  
 ACCESSION U89259  
 NID 91881675  
 KEYWORDS  
 SOURCE Oxytricha fallax  
 ORGANISM Oxytricha fallax  
 Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;  
 hypotrichs; Stichotrichida; Oxytricha.  
 REFERENCE 1 (bases 1 to 354)  
 AUTHORS Doak T.G., Doerder F.P., Jahn C.L. and Herrick G.  
 TITLE A proposed superfamily of transposase genes: transposon-like  
 elements in ciliated protozoa and a common 'D35E' motif  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)  
 MEDLINE 94134747  
 REFERENCE 2 (bases 1 to 354)  
 AUTHORS Witherspoon D.J., Doak T.G., Williams K., Seger J. and Herrick G.  
 TITLE Selection on the protein-coding genes of the TBE1 family of  
 transposable elements in the ciliates Oxytricha fallax and O.  
 trifallax  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 354)  
 AUTHORS Doak T.G., Williams K., Witherspoon D.J. and Herrick G.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah,  
 School of Med. Rm5C334, USA, UT 84132, USA  
 FEATURES  
 source  
 1..354  
 /organism="Oxytricha fallax"  
 /strain="9D1"  
 /transposon="TBE1"  
 /note="this is a bulk sequence that was generated from a  
 PCR product that represents many transposon templates"  
 <1..>354  
 /codon\_start=1  
 /product="57kd zinc finger/protein chimera"  
 /db\_xref="PID:gl881676"  
 /transl\_table=6  
 /translation="HTRDLXKHLKRAHKKXXEXXXXXXKXLRKAREXXXXXX  
 QAXEXXXVXXRNXLSEXTKIMIKIQYKIPVLAQIDLTSLQSYLLEDSFDPKVI  
 XDHQYAKVYFPNYL"  
 BASE COUNT 106 a 42 c 41 g 54 t 111 others  
 ORIGIN  
 Query Match 3.6%; Score 28; DB 11; Length 354;  
 Best Local Similarity 26.6%; Pred. No. 7.55e-03;  
 Matches 37; Conservative 50; Mismatches 50; Indels 2; Gaps 2;  
 Db 30 aaaaagcayaraagadtagarrttgarrakhawgmhwsarytraagytdctcaa 89  
 QY 650 AAGAACCAAAAGATGAGAGTCCCGGAAAATGTTAAAGCATTTCAGTGAAGTATATCTA 709  
 Db 90 haaagaagaar-gcdagrgaryhyhdwargwb-gydgwgcyykaayaagcwagmgarwya 147  
 QY 710 TNNGTATTTTGCAAAATCATTTGTGAACAGTCCACATNTGCTTTTAAACATAGTGTACA 769  
 Db 148 swrtrygtaargawdkrc 166  
 QY 770 ATATTTAGAAAGTTTGAGC 788

RESULT 8  
 LOCUS  
 DEFINITION Rattus myosin heavy chain  
 ACCESSION X62859 S98128  
 NID 957632  
 KEYWORDS actin-activated ATPase activity; myosin II; neuronal myosin heavy  
 chain.

black rat.  
Rattus rattus  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
Murinae; Rattus.  
1 (bases 1 to 6810)  
Chantler, P.D.  
Direct Submission  
Submitted (08-OCT-1991) P.D. Chantler, Prof in Anatomy &  
Neurobiology, M C P, 3200 Henry Avenue, Philadelphia, Pennsylvania  
19129, USA  
2 (bases 1 to 6810)  
Sun, W. and Chantler, P.D.  
Cloning of the cDNA encoding a neuronal myosin heavy chain from  
mammalian brain and its differential expression within the central  
nervous system  
J. Mol. Biol. 224 (4), 1185-1193 (1992)  
92235856  
Location/Qualifiers  
1..6810  
/dev\_stage="adult"  
/organism="Rattus rattus"  
/tissue\_type="brain"  
/cell\_type="neuron"  
/clone\_lib="lambda ZAP cDNA library (Stratagene)"  
1..534  
1..6810  
/evidence="experimental"  
535..6534  
/codon\_start=1  
/product="neuronal myosin heavy chain"  
/db\_xref="PID:g57633"  
/translation="MAQRADKYLVDKNIINPLTQADWAARKLVVVPSEKGFEEA  
SLKEEYGEAIVELAEAGKVKVKKDDIOKMPKFKSVEDMAELCLNASEVYLHLK  
ERYSLGIVTSLGCVINPYKNLPTIPEIVEMVKGKHKHMPHIVAITTAVRS  
MMQDREOSILCTGSGAGKTEKTKVQYLAHVASHSKKQGLERQLQANFIL  
EAFGNATKVNDSNRFGRIFNFDVNGVIGVANIETYLLEKSRALQKAERTPHI  
FYLLSGAGEHLKDLLLVEPKYRFLNSGHVTPQGDQKDFQETMEARIMGIDPE  
EQGLKLVGIVLQNLNIPYKERNTDQASMPDNTAAQKVSHLLNGINVTDFTRGILTP  
RIKVGDIYQKQADTKQADFAI EALAKATYQOMRWLVRINKALDKTKRQAGSFTGI  
LDIAGFEIPLNSFEOLCINVTQKQLQFNHTFLEQVEYREGIEMWFI DFLGLDQ  
PCIDILERPANPGVIALDLDEECWFFPKATDKTFVEKLVQEQGTHSFQPKRQKADK  
FCIIHAGKVDYQKMDPLNDNATLLHQSDSKFEVLEKMDVDRIVGLDQVT  
GTLITAFAGKVTGKGMFRVGLYKESLTKMLATLNTNPNFVCIIPNHEKRAKGL  
DHLVDLQRCNGIRICRGQFPNRIVQEFORYEITLTPNAIPKGFMDGKQACE  
MRALRELDPNLYRIGOSKIFFRAGVLHLEERDLKITDIIIFFOACRGYLAKEAF  
AKQOOLSALKILQRCAVYLKLRHQMWFVTKVPLQVTRQELQAKDELKMK  
VKQTKVLEMEERKHOQLLEKILAEQLQELTFAEEMEARLAKKAEQLEEE  
ILHDSREVEEERKQI LQNEKKQGHKNDLEQLEDEMESARQLQLEKVTTEAKL  
KLEEQIILDEQNLKAKKLEEDRIAEFTNLEEEKSKLAKLNKHEAMITD  
LERLAREKQRELEKTRKLEEDSDLSQDIAELQAOIAELKMLAKKEELQAL  
ARVEEAAQNMALKIRELESIOSELQPLESERASRNKAERKQKRDGLGEALATE  
LEDLTSTAQQLRLEQEVNILKTLEEAKEHQAQIENQKHSQAVELASQL  
EQTKRVANLEKAKQLEDEKNEVLLQGGDESEHKKRVQAEQLEQVLENE  
GERVLTADKQTLVEDLNVTGLLSQSDSKSKLTKDFSALESLOQDTQELLQEN  
RQKLSLTKLVQVEDNSFREQLEEEEAHKNLEKQIATLHAQVADMKMEDSVG  
CLTEAEKQKQKQLEDEKQLEESQHEEKVAAVDKLEKTKRQOELDLLVLDHORSA  
CINLEKQKQKQKQLEDEKQLEESQHEEKVAAVDKLEKTKRQOELDLLVLDHORSA  
FLRNLQEMTERLEDESKVDDAKSVLEHKSILGREQQVMEKTKQLLEDESLQTE  
DAKRLVNLQAKQAFERDQDQDSEKQKLVQVREMEAELEDDQKMSRARA  
AVKLEMDLQDLBAHDSANKNDEAKILRNLAQKQKCMRELDTPRASREIALQA  
KENEKLEMEAMQIOLQEELAAERAKQAOQOERDELADEISNAGSKAGLAKELRL  
LEARIAQLEEELEEGQNTLINDRLKQNLQIDQINADLNLRHQAQKVENARQOLE  
RONKELVKLOEMEVKSKYKASITALEKIAQLEELQDLNTERKQACQKQVRRTEK  
KDLVLQVDDDERNAEQYKQADKASTRLKQKLEERAEAEAAQANRRKRLQET  
EDATEADANNRVSLNKLGRGLDFPVYRLVRKTRLELSDSDDDDDSKASLINETQ  
PQCLQDQDLQFLHWPVNCVCGWGEQVQGEAEVHKCRT"  
535..6534  
/product="MHC head region"  
3037..6315  
/product="MHC rod region"  
6316..6531  
/product="tail piece"

3'UTR 6532..6810  
BASE COUNT 1771 a 1679 c 2165 g 1195 t  
ORIGIN  
Query Match 3.6%; Score 28; DB 92; Length 6810;  
Best Local Similarity 74.18; Pred. No. 7.55e-03;  
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
Db 3954 ggaggccctgaagacagagctggaggagcctgacagactcacagctccccagcaggag 4011  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 495 GGAAGCACTGAAGATGACGATGACCTTCTGACCTCACTGCTCTCCCGCAGGAG 552  
RESULT 9  
LOCUS A10158 201 bp DNA PAT 01-DEC-1993  
DEFINITION S.griseus gene for preprotease (partial).  
ACCESSION A10158  
NID 9490196  
KEYWORDS Streptomyces griseus.  
SOURCE Streptomyces griseus  
ORGANISM Streptomyces griseus  
Eubacteria; Firmicutes; Actinomycetes; Streptomyces;  
Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 201)  
AUTHORS Garvin, R.T. and James, E.  
TITLE Production of active proteins containing cystine residues  
JOURNAL Patent: EP 022279-A 2 20-MAY-1987;  
Cangene Corporation  
FEATURES  
source Location/Qualifiers  
1..201  
/organism="Streptomyces griseus"  
1..>201  
CDS  
/codon\_start=1  
/product="preprotease"  
/db\_xref="PID:g490197"  
/translation="MPHSPVPAESPAPOGPRRPVSRRLLEGGAAVLGALALSAS  
LTAQAAVRAAADEPPEWDFGV"  
BASE COUNT 13 a 66 c 54 g 22 t 46 others  
ORIGIN  
Query Match 3.4%; Score 27; DB 53; Length 201;  
Best Local Similarity 36.9%; Pred. No. 3.16e-02;  
Matches 31; Conservative 24; Mismatches 29; Indels 0; Gaps 0;  
Db 23 csgsgagtcscgscgscgscgscgscgscgscgscgscgscgscgscgscgsc 82  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 272 CGAGCAGGANGCCGCCCTCTGCCCGCTCAAGCGCGGCGACTTACCCCCCGCAGC 331  
Db 83 tsaggsgggsg 106  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 332 TCGGCGCTTCGAGCGGCTCCAGG 355  
RESULT 10  
LOCUS A10161 201 bp DNA PAT 01-DEC-1993  
DEFINITION Synthetic DNA for preprotease leader & prochymosin.  
ACCESSION A10161  
NID 9490655  
KEYWORDS unidentified.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 201)  
AUTHORS Garvin, R.T. and James, E.  
TITLE Production of active proteins containing cystine residues  
JOURNAL Patent: EP 022279-A 5 20-MAY-1987;  
Cangene Corporation  
FEATURES  
source Location/Qualifiers  
1..201  
/organism="Artificial sequences"  
1..>201  
CDS  
/codon\_start=1

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
FEATURES  
source  
5'UTR  
mRNA  
CDS  
misc\_feature  
misc\_feature  
misc\_feature



```

/transl_table=4
/translation="MNNYPLFWINEILNGFVEYLDIFSIMAFUTGIYVILTKNPI
VSVLFLILFGISSYLIIGLNFGLSYIIVIGAVSILFLILMLINRSELSQSN
TSNSPLRTIFIGIFSFLPLPVDIVMLSNFYNNFSEDFVIDWNINDNNLNLY
NNVLFMSVINDGSDVIDFNHTAIGNIYIYNWLIITASILLAMVGSIVITIKQ
RKI"
BASE COUNT 296 a 89 c 115 g 327 t
ORIGIN
Query Match 3.3%; Score 26; DB 63; Length 827;
Best Local Similarity 74.0%; Pred. No. 1.28e-01;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
- Db 187 acaaatggttttgaatacatattagatatattttcaataatggcttt 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 734 ACAATGATTTGC AAAATACANNATAGATATCTCCATCGAATGCTTT 685
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
LOCUS MIPASEB5 1739 bp DNA PLN 31-MAR-1992
DEFINITION Podospora anserina mitochondrial beta sen-DNA 5' fragment.
ACCESSION X03126
NID 913309
KEYWORDS inverted repeat; ND6 gene; plasmid; recombination junction;
transfer RNA; transfer RNA-Asp; transfer RNA-Ser; transfer RNA-Trp;
transfer RNA-Val.
SOURCE Podospora anserina.
ORGANISM Mitochondrion Podospora anserina
Eukaryotes; mitochondrial eukaryotes; Fungi; Ascomycota;
Euscomycetes; Pyrenomycetes; Sordariales; Sordariaceae; Podospora.
REFERENCE 1 (bases 1 to 1739)
AUTHORS Cummings,D.J., MacNeill,I.A., Domenico,J. and Matsuura,E.T.
TITLE Excision-amplification of mitochondrial DNA during senescence in
Podospora anserina. DNA sequence analysis of three unique
plasmids.
J. Mol. Biol. 185 (4), 659-680 (1985)
JOURNAL 86037239
MEDLINE
COMMENT The 54 bp units and the tRNA structures may play a role as
recognition signals for excision. Direct repeat 1 could be involved
in site-specific recombination.
FEATURES
source
Location/Qualifiers
1..1739
/organism="Podospora anserina"
/mitochondrion
misc_feature 23..24
repeat_region /note="J1 junction (excision site of beta sen DNA)"
24..32
/misc_feature 24..>1739
/note="direct repeat 1"
repeat_region /note="beta sen DNA"
45..60
/misc_feature 55..108
/note="5 AGG repeats"
misc_feature /note="54 bp unit"
74..99
/misc_feature /note="imp. palindrome"
162..234
tRNA /note="trNA-Asp"
243..324
tRNA /note="trNA-Ser 2"
364..439
tRNA /note="trNA-Trp"
627..1300
misc_feature /note="ND6 gene"
1340..1393
misc_feature /note="54 bp unit (inverted complement)"
1405..1423
misc_feature /note="palindrome"
1435..1509
tRNA /note="trNA-Val"
1655..1666
misc_feature /note="palindrome"
569 a 263 c 308 g 599 t
BASE COUNT

```

## ORIGIN

```

Query Match 3.3%; Score 26; DB 63; Length 1739;
Best Local Similarity 74.0%; Pred. No. 1.28e-01;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
- Db 669 acaaatggttttgaatacatattagatatattttcaataatggcttt 718
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 734 ACAATGATTTGC AAAATACANNATAGATATCTCCATCGAATGCTTT 685
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
LOCUS DROEAG 4061 bp mRNA INV 12-JUL-1991
DEFINITION D.melanogaster putative potassium channel protein (eag) mRNA,
complete cds.
ACCESSION M61157
NID g157311
KEYWORDS potassium channel protein; transmembrane protein.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 4061)
AUTHORS Warmke,J., Drysdale,R. and Ganetzky,B.
TITLE A distinct potassium channel polypeptide encoded by the Drosophila
eag locus
JOURNAL Science 252, 1560-1562 (1991)
MEDLINE 91262635
FEATURES
source
Location/Qualifiers
1..4061
/organism="Drosophila melanogaster"
/dev_stage="adult"
/tissue_type="head"
464..3988
/feature="eag"
/note="Description: ether a go-go; putative; putative"
/codon_start=1
/product="potassium channel protein"
/db_xref="PID:g157312"
/translation="MPGRRGLVAPONTFLEIRRSNPSQDSFLANAIQVDFPIV
YCNSEFKISGYNRAEVMQSKRYGCMYGTETKRTVGRLEYLTENQDQDFEILL
YKNNLQCCGALSQFGKAOQTETPLWLLQVAPIRNERDLVFLTLTFRITALKQPI
DSEDTKGLSLKFAKLARSVTRSFSAHLPTLKDPTKOSLNLAHMSLSADIMPOYR
QEAPTPPHLLHYCAFKALNDWVILCLFYTAIYVYNVAFNKNKTSSEYSLVDSI
VDVIFIDIVNFHTTFVGPGEVSDPKVIRNMLKSWFILLDLSCLPDVFNAFDR
DEDGIGLSFALKVRLRLGRVVRKLDRLLEYGAAMLILLCLCYMLVAHMLACIWT
IGRSDADNGIQYSLWKLANTQSPYSIWSNDTGPVLVNGSPKRSKMYVYALYFTTC
MTSVGFGNVAATDNEKFTICMIIAALLYATIFGHVTTIQOMTSATKYHDLMLN
VREFMKLHEVPKALSERVMDYVYSTWNTKGLDTEKVLNCCPKDMKADICVHLNRKF
DEHPTFLASDGCCLRALMHFMSHAPGDLTYHTGESIDSLCFYVTSLEIVQDDEV
VAILKGDFVGDQFWDKSAVQSAANRVALTYCDLHAIRKDLLEVLDFEYSAFANSA
RNLYLTYNLRHLIFRVDVYKREKELAEKRNKPEQLPQNODHLVRKIFSKFRTPQV
QAGSKELVGGSGDVEKGDGEVETKVLPAKPLQASQATLARQDTIDGGEVDSSP
PSRDSRVIEGAASVATVGPSPVATTSAAAGAGVSGGSGGTVAIVTKADRNLI
ALERERQIEMASRATSDTYDGLRETPTLAORDLVATVLDKVDVLELRMQOR
IGRIEDLLGELVARLAPGASSGNAPDSSGQTPGDEICAGCGAGGGGTPTQAPT
SAVTSPTVDTITISSPGAGSGGTGAGAGAVAGAGAGGLLDPGATVSSAGGNGLG
PLMKRRRSKSGKAPAPPEQTLASTAGTATAAGVAGSGMTSSAPASDAQQHQHQA
ADQSPPTPGAELHLLEEDFTAALPSTSSGGAGGGGGSGGSGATPTTPTPTIAGGS
GSGTPTSTATTPTTGTGATRGKLDLF"
BASE COUNT 968 a 1105 c 1216 g 772 t
ORIGIN
Query Match 3.3%; Score 26; DB 41; Length 4061;
Best Local Similarity 77.1%; Pred. No. 1.28e-01;
Matches 37; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
- Db 2921 tcgcgcagtggtgtaccacatctcgcgcgcgagctggagta 2968
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 134 TCGCGCCAGTTCGCCACACATCTCGCGAGCCATCTCTCGGAGTA 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



Search completed: Fri Nov 7 14:55:41 1997  
Job time : 710 secs.



```

PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI; 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B. NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match      8.7%; Score 68; DB 2; Length 1047;
Best Local Similarity 7.3%; Pred. No. 9.65e-27;
Matches 41; Conservative 166; Mismatches 350; Indels 4; Gaps 4;

Db 345 avnnntnnngtrndgnrnvnmkngrryhgvtgnvnmkndrntdnvnmwngdndsgd 404
Qy 230 AGATCGTGGCGGGACCAAGCGCGCGCCGACGACGACGACGACGACGACGACGCGCC 289
Db 405 nnaahysganknnwtgrnnnwvkgannsdnnncandndnscktnstnanvngtg 464
Qy 290 CTCCTGCCCGCTCAAGCGGCG-CGACTTCACCCCGCGGAGCTGGCGGCTTCGACGGC 348
Db 465 ntnnmgvssnnnrknmnnknnaasmwrrnnnnnnngnsrnyhkgagsrntnsrgssy 524
Qy 349 GTCCAGGACCGCGCATACTCATGCCCATCAACGCGCAAGGTGTCTGATGTGACCAAGGC 408
Db 525 gsnmtahgkynnantghnkvgnvankhvknnrnntrnnvnnnkhmrdvnnnhntrng 584
Qy 409 CGCAAAATTCACGGGCGCGGAGGGCGGTATGGGGTCTTTGCTGGAAGAGATCCACAGG 468
Db 585 acndnnnncvntnrcngsnndnnnnndnnndvnnrnsnnndvkmannhnsnshgs 644
Qy 469 GGCCCTTGCCACATTTTGCTCGTAAAGAGACACTGAAGGATGAGTACGATGACCTTCT 528
Db 645 nksscvcvdsrvnknktdvgnasrntandnnanyaknnntannnnsgnnnttgma 704
Qy 529 GACCTCAGTGTGCCAGCAGGAGACTCTCAGTACTGGAGT-CTCAGTTCACTTCAA 587
Db 705 advysngnnnnnnaansrnyngndnsknvknvknvngnnyrnnsndrtnnnnnvnn 764
Qy 588 GTATCATCACGTG-GGCAAACTGCTGAAGGAGGGGAGGAGCCACTGTGTACTCAGATG 646
Db 765 mnrvcvandanrdngnknrrnnknngtsnnndnnnnnnnnnnnnnnnnnnrtayn 824
Qy 647 AGG-AAGAAACCAAAAGATGAGAGTTCCTCCGGAAAAATGTTAAAGCATTCAGTGAAGTATA 705
Db 825 nkrkanannynnnhsvannnkrgntvnanandsvtnysndvngtansanmnvvt 884
Qy 706 TCTATNTGTATTTTCCAAAATCATTTGTAAACAGTCCACTNTGCTCTTAAACATAGTGT 765
Db 885 nnndhytcdandndndvykv 905
Qy 766 TACAATATTTAGAAAAGTTGA 786

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.

```

```

FH Peptide Location/Qualifiers
FT 1..22 /label= signal sequence
FT 12 /label= mature NPRB
FT 23..455 /label= extracellular domain
FT /label= "binds natriuretic peptides A,B and C]"
FT 456..456 /label= transmembrane domain
FT 479..1047 /label= cytoplasmic domain
FT /label= "GC and protein kinase activity"
FT Modified-site 24..26 /label= N-glycos_site
FT Modified-site 35..37 /label= N-glycos_site
FT Modified-site 161..163 /label= N-glycos_site
FT Modified-site 195..197 /label= N-glycos_site
FT Modified-site 244..246 /label= N-glycos_site
FT Modified-site 277..279 /label= N-glycos_site
FT Modified-site 349..351 /label= N-glycos_site
FT Modified-site 600..602 /label= N-glycos_site
PN W09100292-A.
PD 10-JAN-1991.
PD 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI; 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B. NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match      7.7%; Score 60; DB 2; Length 1047;
Best Local Similarity 8.4%; Pred. No. 1.06e-21;
Matches 41; Conservative 138; Mismatches 304; Indels 3; Gaps 3;

Db 434 sdnnncandndnscktnstnanvngtgnnnngvssnnnnrknmnnknnaasmwr 493
Cp 495 CTTATCAGGCAAAATGTGCAAGGCCCTTGATGATCTCTTCAGCAAGAGACCCATA 436
Db 494 nrwnnnnnnnsrnyhkgagsrntnsrgssyngsmntahgkynnantghnkvgnvankhv 553
Cp 435 CGGCCCTCGGGCGCGTAGAAATTTGGCGGCTTTGGTCACATCGAACACCTTGCGGTGAT 376
Db 554 nkrrnntrnnvnnnkhmrdvnnnhntrngacndnnnnnnnnnnnnnnnnnnndn 613
Cp 375 GGCCATGAGTATCGCGGGTCTCTGGAGCGCTCGAAGCGCGGAGCTCGCGGGGG-TGA 317
Db 614 nndwmrnyssnnndvkgmannhnsnshgsnksncvvdsvrnnvknktdvgnasnrtan 673
Cp 316 AGTCGGCGCGCTTGAGCGGGGGGAGAGGGGGCGGCGTCGTCGTCCTGTCGCCGCTGG 257
Db 674 nddnnanyaknnntannnnnnnnnnnttgmnaadvysngnnnnnnnnnnnnnnnnndn 733

```

|                       |   |
|-----------------------|---|
| CC                    | Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14 |
| CC                    | (Q51735). It hybridized to all spp. of mycobacteria tested, but   |
| CC                    | Cross reacted to a few non-mycobacterial spp. The probe may       |
| CC                    | be useful as an initial screen for mycobacterial infection.       |
| CC                    | See also Q51735-45 and Q51747-59.                                 |
| SC                    | Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;                             |
| Query Match           | 5.5%; Score 43; DB 9; Length 91;                                  |
| Best Local Similarity | 10.5%; Pred. No. 2.20e-11;  |
| Matches               | 6; Conservative 44; Mismatches 7; Indels 0; Gap                   |
| Db                    | 5 ccggcgagshvyyvvhhsvhhsvhvvhvhsvvvvhhvvhvvhvhyrvsvvc 61          |
|                       | :         |
| Cp                    | 158 CGCGCGCTTCCAGATCGTTGGTCGCGGCAGTCGCCACCATCCTTCGCAGCC 10        |

|        |   |
|--------|---|
| RESULT | 5   |
| ID     | N81164 standard; DNA; 204 BP.   |
| AC     | N81164;   |
| DE     | 8-NOV-1990 (first entry)  |
| DT     | Base substituted E.coli beta-galactosidase alpha-fragment.            |
| KW     | E.coli beta galactosidase alpha-fragment; base substitutions; ss.     |
| OS     | Escherichia coli.   |
| FH     | Key Location/Qualifiers   |
| FT     | misc_feature 19..69   |
| FT     | *tag= a   |
| FT     | /function-multiple cloning site                                       |
| FT     | primer_bind 187..204  |
| FT     | *tag= b   |
| PN     | EP-285123-A.  |
| PD     | 05-MAY-1988.  |
| PP     | 30-MAR-1988; 105163.  |
| PR     | 03-APR-1987; US-034819.   |
| PA     | (SUSO) SUOMEN SOKERI OY.  |
| PB     | Lehetoara P, Knowles J, Koivula A, Bamford J, Reinikainen T;          |
| DR     | WPI; 88-279927/40.  |
| PT     | Introducing random point mutations into nucleic acids -               |
| PT     | by prepn of single stranded template, annealing a primer, elongation, |
| PT     | misincorporation, completion of molecules and screening.              |
| PT     | Disclosure; p; English.   |
| PC     | Random point mutations were introduced into the alpha fragment of     |
| CC     | E.coli beta-galactosidase. The wild type sequence was obtained as a   |
| CC     | single stranded template and an oligonucleotide was hybridised to     |
| CC     | it to generate a popn of DNA molecules which terminate at all         |
| CC     | possible nucleotide positions within a specified region. The          |
| CC     | variable 3' ends generated in this way are used as primers for        |
| CC     | reverse transcriptase. Nucleotides are misincorporated by the         |
| CC     | transcriptase and the molecules are completed to forms that can be    |
| CC     | amplified and then expressed in a suitable host-vector system.        |
| CC     | The sequence covers all 176 diff base substitutions, most of which    |
| CC     | occurred singularly in any given mutant.                              |
| CC     | See also P80575.  |
| CC     | Sequence 204 BP.  |
| SQ     | 21 A; 47 C; 17 G; 11 T; 108 Others;                                   |

|                 |   |  |              |               |                               |
|-----------------|---|--|--------------|---------------|-------------------------------|
|                 | Query Match   | 5.4%   | Score 42;    | DB 1;         | Length 204;                   |
|                 | Best Local Similarity   | 14.3%  | Pred.        | No. 8.44e-11; |                               |
|                 | Matches   | 17;  | Conservative | 56;           | Mismatches 46; Indels 0; Gaps |
| D <b>b</b>      | 74  | aycydcgvscgcmrththhyrrmbvrvdyndrsdaawwyccyrrsvkydvccynachhd  | 1333         |               |                               |
| Q <b>Y</b>      | 172   | ATTTTCACGTGGCGGCTCAACCTGCTGCCTGGCCTCTGCAFCATCTCGTCTACAAG     | 231          |               |                               |
| D <b>b</b>      | 134   | dhyvybbbyvvnhnancnccebnhnchvnhvbnnhrnwaiyvzhdarrddvhcevchccc | 192          |               |                               |
| Q <b>Y</b>      | 232   | ATCGTGCGGGGGACCAGCCGCCGGCCAGCGGCGGACAGGACGACGACGANGCGCGCCCC  | 290          |               |                               |
| R <b>E</b> SULT | 6   |  |              |               |                               |
| I <b>C</b>      | N81164  | standard; DNA;   | 204 BP.      |               |                               |
| A <b>D</b>      | N81164;   |  |              |               |                               |
| D <b>T</b>      | 08-NOV-1990   | (first entry)  |              |               |                               |
| K <b>E</b>      | Base substituted E.coli beta-galactosidase alpha-fragment.        |  |              |               |                               |
| D <b>W</b>      | E.coli beta galactosidase alpha-fragment; base substitutions; ss. |  |              |               |                               |

```

OS Escherichia coli.
FH key Location/Qualifiers
FT misc_feature 19..69
FT /*tag= a
FT /function-multiple cloning site
FT primer_bind 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PR 30-MAR-1988. 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT minicorporation, completion of molecules and screening.
PS Disclosure: P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also F80375.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 5.4%; Score 42; DB 1; Length 204;
Best Local Similarity 17.5%; Pred. No. 8.44e-11;
Matches 28; Conservative 68; Mismatches 63; Indels 1; Gaps 1;

Db 30 yrcbpcaggygacbcyrraggnycccggywccgagcyogaaycdhgcgymrt 89
: : | | | | | : : : : : | | | | | | | : : : : :
Cp 276 GTCGCTCTCCATCGCTGGTGGCGCGCGCTGGTCCCGCGCAGCATCTGTAGACAGGAA 217
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 tthyrnrbnrvdyrnrsdaaayccyrsvkydc-ynachddhdyvbbvynvnh 148
: : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 216 GATCAGAGGCGCAAGCAGCAGCAGGTGAGCGCGCAGTGAAATCTCATGCAGCAGCC 157
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 nncnccbnhchvbnhbnhnrwayvrhddrddvhccv 188
: : | : : : : : : : : : : : : : : : : : : : : : : :
Cp 156 GCCGCTCTCCATCGCTGGTGGCGCGCGCAGTCGCCACC 117

RESULT 7
ID Q70468 standard; DNA; 114 BP.
AC Q70468;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'z'; z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be

```

```

DR P-PSDB; R65154.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)11(TGC)(NNB)62(NNB)7(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed activity
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 4.8%; Score 38; DB 12; Length 114;
Best Local Similarity 5.4%; Pred. No. 1.69e-08;
Matches 6; Conservative 34; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn 62
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 GTCCGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 239
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn 114
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 CGGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCCCT 291
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'z'; z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be

```





[illegible]



Search completed: Fri Nov 7 14:57:38 1997  
Job time : 97 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Fri Nov 7 15:31:42 1997; MasPar time 37.13 Seconds  
920.374 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-822-264-2  
Description: (J-788) from US08822264.seq  
Perfect Score: 784

N.A. Sequence: 1 GCGGCGGAACCGCGCGCC.....AATATTAGAAAGTTTGAGC 788  
Comp: CGGCGCGTTGGGCGCGCGG.....TTATAATCTTTCAAACTCG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 82050 seqs, 21682983 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-issued  
1:back1 2:51 3:52 4:53 5:54 6:55 7:56 8:PT90 9:PT91  
10:PT92 11:PT93 12:PT94 13:PT95 14:PT96

Statistics: Mean 8.477; Variance 4.723; scale 1.795

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                        | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------------------|-----------|
| C 1        | 53    | 6.8         | 7218   | 7     | US-08-232- Sequence 14, Applicatio | 8.00e-20  |
| C 2        | 40    | 5.1         | 215    | 6     | US-08-238- Sequence 5, Applicatio  | 2.11e-11  |
| C 3        | 35    | 4.5         | 7218   | 7     | US-08-232- Sequence 14, Applicatio | 2.58e-08  |
| C 4        | 31    | 4.0         | 215    | 6     | US-08-238- Sequence 5, Applicatio  | 6.25e-06  |
| C 5        | 26    | 3.3         | 74     | 13    | PCT-US95-1 Sequence 94, Applicatio | 4.32e-03  |
| C 6        | 26    | 3.3         | 81     | 13    | PCT-US95-1 Sequence 92, Applicatio | 4.32e-03  |
| C 7        | 24    | 3.1         | 74     | 13    | PCT-US95-1 Sequence 100, Applicat  | 5.18e-02  |
| C 8        | 24    | 3.1         | 75     | 13    | PCT-US95-1 Sequence 99, Applicatio | 5.18e-02  |
| C 9        | 24    | 3.1         | 81     | 13    | PCT-US95-1 Sequence 98, Applicatio | 5.18e-02  |
| C 10       | 24    | 3.1         | 82     | 13    | PCT-US95-1 Sequence 97, Applicatio | 5.18e-02  |
| C 11       | 24    | 3.1         | 82     | 13    | PCT-US95-1 Sequence 97, Applicatio | 5.18e-02  |
| C 12       | 23    | 2.9         | 74     | 13    | PCT-US95-1 Sequence 100, Applicat  | 1.74e-01  |
| C 13       | 23    | 2.9         | 74     | 13    | PCT-US95-1 Sequence 94, Applicatio | 1.74e-01  |
| C 14       | 23    | 2.9         | 81     | 13    | PCT-US95-1 Sequence 98, Applicatio | 1.74e-01  |
| C 15       | 23    | 2.9         | 81     | 13    | PCT-US95-1 Sequence 92, Applicatio | 1.74e-01  |
| C 16       | 23    | 2.9         | 81     | 13    | PCT-US95-1 Sequence 92, Applicatio | 1.74e-01  |
| C 17       | 22    | 2.8         | 66     | 7     | US-08-471- Sequence 144, Applicat  | 5.68e-01  |
| C 18       | 22    | 2.8         | 66     | 7     | US-08-471- Sequence 144, Applicat  | 5.68e-01  |
| C 19       | 22    | 2.8         | 69     | 7     | US-08-471- Sequence 142, Applicat  | 5.68e-01  |

|   |    |    |     |       |    |                                   |          |
|---|----|----|-----|-------|----|-----------------------------------|----------|
| C | 20 | 22 | 2.8 | 69    | 7  | US-08-471- Sequence 142, Applicat | 5.68e-01 |
|   | 21 | 22 | 2.8 | 84    | 7  | US-08-133- Sequence 120, Applicat | 5.68e-01 |
|   | 22 | 22 | 2.8 | 84    | 7  | US-08-300- Sequence 25, Applicati | 5.68e-01 |
|   | 23 | 22 | 2.8 | 84    | 6  | US-08-209- Sequence 23, Applicati | 5.68e-01 |
|   | 24 | 22 | 2.8 | 84    | 13 | PCT-US95-1 Sequence 25, Applicati | 5.68e-01 |
|   | 25 | 22 | 2.8 | 84    | 12 | PCT-US94-0 Sequence 25, Applicati | 5.68e-01 |
| C | 26 | 22 | 2.8 | 84    | 6  | US-08-209- Sequence 23, Applicati | 5.68e-01 |
|   | 27 | 22 | 2.8 | 84    | 7  | US-08-300- Sequence 25, Applicati | 5.68e-01 |
| C | 28 | 22 | 2.8 | 84    | 13 | PCT-US95-1 Sequence 25, Applicati | 5.68e-01 |
|   | 29 | 22 | 2.8 | 84    | 12 | PCT-US94-0 Sequence 25, Applicati | 5.68e-01 |
| C | 30 | 22 | 2.8 | 84    | 7  | US-08-133- Sequence 120, Applicat | 5.68e-01 |
|   | 31 | 22 | 2.8 | 2334  | 13 | PCT-US95-0 Sequence 1, Applicatio | 5.68e-01 |
| C | 32 | 21 | 2.7 | 56    | 13 | PCT-US95-1 Sequence 87, Applicati | 1.81e+00 |
|   | 33 | 21 | 2.7 | 65    | 7  | US-08-471- Sequence 145, Applicat | 1.81e+00 |
| C | 34 | 21 | 2.7 | 68    | 7  | US-08-471- Sequence 143, Applicat | 1.81e+00 |
|   | 35 | 21 | 2.7 | 85    | 12 | PCT-US94-0 Sequence 1, Applicatio | 1.81e+00 |
|   | 36 | 21 | 2.7 | 85    | 7  | US-08-370- Sequence 26, Applicati | 1.81e+00 |
|   | 37 | 21 | 2.7 | 85    | 12 | PCT-US94-0 Sequence 26, Applicati | 1.81e+00 |
| C | 38 | 21 | 2.7 | 242   | 7  | US-08-273- Sequence 1, Applicatio | 1.81e+00 |
|   | 39 | 21 | 2.7 | 1794  | 12 | PCT-US94-1 Sequence 9, Applicatio | 1.81e+00 |
| C | 40 | 20 | 2.6 | 66    | 13 | PCT-US95-1 Sequence 93, Applicati | 5.59e+00 |
|   | 41 | 20 | 2.6 | 786   | 13 | PCT-US95-0 Sequence 1, Applicati  | 5.59e+00 |
| C | 42 | 20 | 2.6 | 2244  | 13 | PCT-US95-0 Sequence 10, Applicati | 5.59e+00 |
|   | 43 | 20 | 2.6 | 5125  | 14 | PCT-US96-0 Sequence 4, Applicatio | 5.59e+00 |
| C | 44 | 20 | 2.6 | 15664 | 7  | US-08-402- Sequence 3, Applicatio | 5.59e+00 |
|   | 45 | 20 | 2.6 | 15664 | 6  | US-08-508- Sequence 3, Applicatio | 5.59e+00 |

## ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DT 01-JAN-1900  
DE Sequence 14, Application US/08232463.  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)683-4109  
CC TELEX: 899149  
CC INFORMATION FOR SEQ ID NO: 14:





```

CC MOLECULE TYPE: DNA (genomic)
SQ Sequence 74 BP; 3 A; 4 C; 1 G; 1 T; 63 other;
    Query Match 3.3%; Score 26; DB 13; Length 74;
    Best Local Similarity 8.7%; Pred. No. 4.32e-03;
    Matches 6; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

Db 2 CGAGNNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBN 61
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 278 CGAGCGCGCCCTCTGCCCGCGCTCAAGCGGGCGGCGACTTCACCCCGCGGAGCTCGGC 337
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 NBNNBNACG 70
    : : |||
Qy 338 GCTTCGACG 346

RESULT 6
ID PCT-US95-11934-92 STANDARD; DNA; UNC; 81 BP.
AC xxxxxx
DT 01-JAN-1900
E Sequence 92, Application PC/TUS9511934.
F Sequence 92, Application PC/TUS9511934
G GENERAL INFORMATION:
CC APPLICANT: Cytogen Corporation
CC TITLE OF INVENTION: Antigen Binding Peptides (Aptides) From
CC TITLE OF INVENTION: Peptide Libraries
CC NUMBER OF SEQUENCES: 103
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/11934
CC FILING DATE: 20-SEP-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Misrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 92:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 81 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
SQ Sequence 81 BP; 3 A; 5 C; 6 G; 4 T; 63 other;

Query Match 3.3%; Score 26; DB 13; Length 81;
Best Local Similarity 8.7%; Pred. No. 4.32e-03;
Matches 6; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

Db 9 CGAGNNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBN 68
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 278 CGAGCGCGCCCTCTGCCCGCGCTCAAGCGGGCGGCGACTTCACCCCGCGGAGCTCGGC 337
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 NBNNBNACG 77
    : : |||
Qy 338 GCTTCGACG 346

```

[illegible]

|        |   |  |
|--------|---|--|
| CC     | TELEPHONE: (212) 790-9090                                   |  |
| CC     | TELEFAX: (212) 869-9741/8864                                |  |
| CC     | TELEX: 66141 PENNIE   |  |
| CC     | INFORMATION FOR SEQ ID NO: 99:                              |  |
| CC     | SEQUENCE CHARACTERISTICS:                                   |  |
| CC     | LENGTH: 75 base pairs                                       |  |
| CC     | TYPE: nucleic acid  |  |
| CC     | STRANDEDNESS: single  |  |
| CC     | TOPOLOGY: linear  |  |
| CC     | MOLECULE TYPE: DNA (genomic)                                |  |
| CC     | SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 other;               |  |
| CC     | Query Match   | 3.1%; Score 24; DB 13; Length 75;                    |
| CC     | Best Local Similarity                                       | 7.7%; Pred. No. 5.18e-02;                            |
| CC     | Matches   | 5; Conservative 20; Mismatches 40; Indels 0; Gaps 0; |
| Db     | 3 GAGNNBN       | 62   |
| Cp     | 292 GAGGGGGCGGNTCGTCTGCTGCTGCTGCGCGCTGGCGCGGCTGGTCCCGGCA    | 233  |
| Db     | 63 BNNBG 67   |  |
| Cp     | 232 TCTTG 228   |  |
| RESULT | 10  |  |
| ID     | PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.                |  |
| AC     | xxxxxx  |  |
| Dt     | 01-JAN-1900   |  |
| DE     | Sequence 98, Application PC/TUS9511934.                     |  |
| CC     | Sequence 98, Application PC/TUS9511934                      |  |
| CC     | GENERAL INFORMATION:  |  |
| CC     | APPLICANT: Cytozen Corporation                              |  |
| CC     | TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From |  |
| CC     | TITLE OF INVENTION: Peptide Libraries                       |  |
| CC     | NUMBER OF SEQUENCES: 103                                    |  |
| CC     | CORRESPONDENCE ADDRESS:                                     |  |
| CC     | ADDRESSEE: Pennie & Edmonds                                 |  |
| CC     | STREET: 1155 Avenue of the Americas                         |  |
| CC     | CITY: New York  |  |
| CC     | STATE: New York   |  |
| CC     | COUNTRY: USA  |  |
| CC     | ZIP: 10036  |  |
| CC     | COMPUTER READABLE FORM:                                     |  |
| CC     | MEDIUM TYPE: Floppy disk                                    |  |
| CC     | COMPUTER: IBM PC compatible                                 |  |
| CC     | OPERATING SYSTEM: PC-DOS/MS-DOS                             |  |
| CC     | SOFTWARE: PatentIn Release #1.0, Version #1.30              |  |
| CC     | CURRENT APPLICATION DATA:                                   |  |
| CC     | APPLICATION NUMBER: PCT/US95/11934                          |  |
| CC     | FILING DATE: 20-SEP-1995                                    |  |
| CC     | CLASSIFICATION:   |  |
| CC     | ATTORNEY/AGENT INFORMATION:                                 |  |
| CC     | NAME: Mirock, S. Leslie                                     |  |
| CC     | REGISTRATION NUMBER: 18,872                                 |  |
| CC     | REFERENCE/DOCKET NUMBER: 1101-196-228                       |  |
| CC     | TELECOMMUNICATION INFORMATION:                              |  |
| CC     | TELEPHONE: (212) 790-9090                                   |  |
| CC     | TELEFAX: (212) 869-9741/8864                                |  |
| CC     | TELEX: 66141 PENNIE   |  |
| CC     | INFORMATION FOR SEQ ID NO: 98:                              |  |
| CC     | SEQUENCE CHARACTERISTICS:                                   |  |
| CC     | LENGTH: 81 base pairs                                       |  |
| CC     | TYPE: nucleic acid  |  |
| CC     | STRANDEDNESS: single  |  |
| CC     | TOPOLOGY: linear  |  |
| CC     | MOLECULE TYPE: DNA (genomic)                                |  |
| CC     | SEQUENCE 81 BP; 6 A; 6 C; 4 G; 5 T; 60 other;               |  |
| CC     | Query Match   | 3.1%; Score 24; DB 13; Length 81;                    |
| CC     | Best Local Similarity                                       | 6.2%; Pred. No. 5.18e-02;                            |
| CC     | Matches   | 4; Conservative 20; Mismatches 40; Indels 0; Gaps 0; |

Query Match 3.1%; Score 24; DB 13; Length 81;  
Best Local Similarity 6.2%; Pred. NO. 5.18e-02;  
Matches 4; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

[illegible]





Search completed: Fri Nov 7 15:45:42 1997  
Job time : 840 secs.



QY 32 LLLGLCI-FLLYKIVRGDQPAASGDRTTTXXPPPLPRL 68

|                       |   |
|-----------------------|---|
| RESULT                | 4   |
| ENTRY                 | A34259 #type fragment   |
| TITLE                 | cytochrome P450mt4, phenobarbital-inducible, mitochondrial,<br>hepatic - rat (fragment)   |
| ORGANISM              | #formal_name Rattus norvegicus #common_name Norway rat  |
| DATE                  | 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change<br>22-Jul-1994  |
| ACCESSIONS            | A34259  |
| REFERENCE             | A34259  |
| #authors              | Shayig, R.M.; Avadhani, N.G.  |
| #journal              | Biochemistry (1990) 29:866-873  |
| #title                | A phenobarbital-inducible hepatic mitochondrial cytochrome<br>P-450 immunochemically related to microsomal P-450b.                |
| #cross-references     | MUID:190254127  |
| #accession            | A34259  |
| #molecule_type        | protein   |
| #residues             | 1-42 ##label SHA  |
| CLASSIFICATION        | #superfamily cytochrome P450  |
| KEYWORDS              | electron transfer; heme; mitochondrion; monooxygenase;<br>oxidoreductase  |
| SUMMARY               | #length 42 #checksum 9226   |
| Query Match           | 6.2%; Score 96; DB 5; Length 42;  |
| Best Local Similarity | 44.7%; Pred.No. 3.23e+00;   |
| Matches               | 17; Conservative 8; Mismatches 9; Indels 4; Gaps 4;   |
| Dn                    | 6 lllallvfglll-lvrg-hpksgn-fpkgprlpkm 40<br>      :      :      :   :   |
| Qy                    | 32 LLLGLGCI-FLLYKIVRGDPAASGDRTTXXPPPLRL 68<br>      :      :      :   :   |
| RESULT                | 5   |
| ENTRY                 | JN0811 #type complete   |
| TITLE                 | benzene 1,2-dioxygenase (EC 1.14.12.3) ferredoxin component<br>Pseudomonas putida plasmid pBMT112                                 |
| ORGANISM              | #formal_name Pseudomonas putida   |
| DATE                  | 19-May-1994 #sequence_revision 06-Jan-1995 #text_change<br>10-Feb-1995  |
| ACCESSIONS            | JN0811; S00559  |
| REFERENCE             | JN0810  |
| #authors              | Tan, H.M.; Tang, H.Y.; Joannou, C.L.; Abdel-Wahab, N.H.;<br>Mason, J.R.   |
| #journal              | Gene (1993) 130:33-39   |
| #title                | The pseudomonas putida ML2 plasmid-encoded genes for benzene<br>dioxygenase are unusual in codon usage and low in G+C<br>content. |
| #accession            | JN0811  |
| #molecule_type        | DNA   |
| #residues             | ##residues 1-107 ##label TAN  |
| REFERENCE             | ##experimental_source strain ML2<br>S00559  |
| #authors              | Morrice, N.; Geary, P.; Cammack, R.; Harris, A.; Beg, F.;<br>Aitken, A.   |
| #journal              | FBS Lett. (1988) 231:336-340  |
| #title                | Primary structure of protein B from Pseudomonas putida,<br>member of a new class of 2Fe-2S ferredoxins.                           |
| #cross-references     | MUID:88196420   |
| #accession            | S00559  |
| #molecule_type        | protein   |
| #residues             | ##residues 2-107 ##label MOR  |
| COMMENT               | This enzyme component reduces the terminal oxygenase in the<br>oxidation of benzene to cis-1,2-dihydroxy-cyclohexa-3,5-diene.     |
| GENETICS              |   |
| #gene                 | bedB  |
| #genome               | plasmid   |
| CLASSIFICATION        | #superfamily toluene dioxygenase ferredoxin component   |
| KEYWORDS              | 2Fe-2S; aromatic hydrocarbon catabolism; iron-sulfur protein;<br>oxidoreductase   |
| FEATURE               | #binding_site 2Fe-2S cluster (Cys) (covalent) #status<br>43..62   |

```

45,65      #binding_site 2Fe-2S cluster (His) (ligands) #status
predicted
SUMMARY    #length 107 #molecular-weight 11940 #checksum 9936

Query Match      6.2%; Score 95; DB 1; Length 107;
Best Local Similarity 30.3%; Pred. No. 4.17e+00;
Matches 10; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

Db 7 lrgsdilppemaryeggsopvmcndvgeffav 39
QY 68 LRRDFTPAELRRFGVQDPRIILMAINGKVFV 100

RESULT 6
ENTRY #J01539 #type complete
TITLE hypothetical 14.5K protein - Salmonella typhimurium plasmid
ORGANISM NTP16
#formal_name Salmonella typhimurium
#sequence_revision 30-Sep-1993 #text_change
#accession J01539
REFERENCE J01538
#authors Cannon, P.M.; Strike, P.
#journal Plasmid (1992) 27:220-230
#title Complete nucleotide sequence and gene organization of plasmid
NTP16.
#accession J01539
#status translation not shown
#molecule_type DNA
#residues 1-126 #label CAN
GENETICS
#genome
#plasmid
SUMMARY #length 126 #molecular-weight 14527 #checksum 6824

Query Match      6.2%; Score 96; DB 9; Length 126;
Best Local Similarity 31.7%; Pred. No. 3.23e+00;
Matches 13; Conservative 16; Mismatches 10; Indels 2; Gaps 2;

Db 60 gkkdry-vytsfngk-fasytlntkvtktdeyndlselsas 98
QY 108 GPEGPYGVFAGRDASRLTCLDKALKDEYDLSLTAA 148

RESULT 7
ENTRY #S57106 #type complete
TITLE probable membrane protein YJR087w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein J1870
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
01-Mar-1996
ACCESSIONS S57106
REFERENCE S57085
#authors Manus, V.; Huang, M.E.; Galibert, F.
#submission submitted to the Protein Sequence Database, September 1995
#accession S57106
#molecule_type DNA
#residues 1-116 #label MAN
#cross-references EMBL:Z49586
GENETICS
#map_position 10R
KEYWORDS transmembrane protein
SUMMARY #length 116 #molecular-weight 13145 #checksum 2626

Query Match      6.1%; Score 94; DB 12; Length 116;
Best Local Similarity 36.4%; Pred. No. 5.38e+00;
Matches 12; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

Db 5 eskkknhafsyplspylfssnfgsvhilfk 37
QY 188 DESSRNKRVAFSGSISXX-YFAKSFVTHVXFK 219

```

```

RESULT 8
ENTRY #B32055 #type complete
TITLE nifv protein - Azotobacter vinelandii
ORGANISM #formal_name Azotobacter vinelandii
DATE 13-Jul-1989 #sequence_revision 03-Aug-1992 #text_change
23-Mar-1993
ACCESSIONS B32055
REFERENCE A32055
#authors Jacobson, M.R.; Brigle, K.E.; Bennett, L.T.; Setterquist, R.A.; Wilson, M.S.; Cash, V.L.; Beynon, J.; Newton, W.E.; Dean, D.R.
#journal J. Bacteriol. (1989) 171:1017-1027
#title Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii.
#cross-references MUID:89123097
#accession B32055
#status preliminary
#molecule_type DNA
#residues 1-242 #label JAC
SUMMARY #length 242 #molecular-weight 26702 #checksum 6919

Query Match      6.1%; Score 94; DB 8; Length 242;
Best Local Similarity 29.7%; Pred. No. 5.38e+00;
Matches 19; Conservative 15; Mismatches 26; Indels 4; Gaps 4;

Db 171 lladcq-llvtlsigppaakvrvagvhpvlralarpareiv-eelqrvtatappwlaak 228
QY 34 LLGLCIFLLYKIVRGDQPAASGDRITTYPPPLRLKR-RDFTPAELRR-FDGVQDPRIILM 91
Db 229 amga 232
QY 92 AING 95

RESULT 9
ENTRY #C36516 #type complete
TITLE toluene dioxygenase (EC 1.14.12.11) ferredoxin component - Pseudomonas putida
ALTERNATE_NAMES 10K benzene oxidation protein; benzene 1,2-dioxygenase ferredoxin component
ORGANISM #formal_name Pseudomonas putida
DATE 15-Feb-1991 #sequence_revision 06-Jan-1995 #text_change
12-Apr-1996
ACCESSIONS C36516; C29830
REFERENCE A36516
#authors Zylstra, G.J.; Gibson, D.T.
#journal J. Biol. Chem. (1989) 264:14940-14946
#title Toluene degradation by Pseudomonas putida F1. Nucleotide sequence of the todC1C2BADE genes and their expression in Escherichia coli.
#cross-references MUID:89359301
#accession C36516
#molecule_type DNA
#residues 1-107 #label ZYL
#cross-references GB:J04996
#experimental_source strain F1
REFERENCE A91848
#authors Irie, S.; Doi, S.; Yorifuji, T.; Takagi, M.; Yano, K.
#journal J. Bacteriol. (1987) 169:5174-5179
#title Nucleotide sequencing and characterization of the genes encoding benzene oxidation enzymes of Pseudomonas putida.
#cross-references MUID:88032840
#accession C29830
#molecule_type DNA
#residues 17-107 #label IRI
#experimental_source strain 136R-3
GENETICS
#gene todB
CLASSIFICATION #superfamily toluene dioxygenase ferredoxin component
KEYWORDS 2Fe-2S; aromatic hydrocarbon catabolism; iron-sulfur protein; oxidoreductase
FEATURE

```

5

```
#accession A28940
#molecule_type DNA
#residues 1-523 #label GRE
REFERENCE A35712
#authors Duewke, T.J.; Gennis, R.B.
#journal J. Biol. Chem. (1990) 265:4273-4277
#title Epitopes of monoclonal antibodies which inhibit ubiquinol
oxidase activity of Escherichia coli cytochrome d complex
localize functional domain.
#cross-references MUID:90170924
#accession A35712
#status preliminary; not compared with conceptual translation
#molecule_type DNA
#residues 253-263 #label DUE
GENETICS
#gene cyd-1
#map_position 17
KEYWORDS membrane protein; oxidoreductase; respiratory chain
SUMMARY #length 523 #molecular-weight 58345 #checksum 1340
Query Match 6.0%; Score 92; DB 9; Length 523;
Best Local Similarity 26.3%; Pred. No. 8.88e+00;
Matches 15; Conservative 15; Mismatches 26; Indels 1; Gaps 1;
Db 453 aigevlptavansltagdlifsmvlicglytlflvaelfmxfarlgpssrktgr 509
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
QY 2 AAEVDVATGADPSDESGLLHE-IFTSPLNLLLLGLCIFLL-YKIVRGDQPAASGDR 57
3' ends.
#cross-references MUID:91348257
#accession S17150
#molecule_type DNA
#residues 1-581 #label LUN
SUMMARY #length 581 #molecular-weight 64036 #checksum 8607
Query Match 6.0%; Score 92; DB 14; Length 581;
Best Local Similarity 32.7%; Pred. No. 8.88e+00;
Matches 34; Conservative 14; Mismatches 51; Indels 5; Gaps 5;
Db 12 invggrthetyrstkltp-gtrllalassepgqdcitaaqdklqplpppspprpppl 70
|:| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 16 LESGGLLHEIFTSPLNLLGLCIFLL-YKIVRGDQPAASGDRTTTXXPPL-PRLKRRDF 73
07-Oct-1994
Db 71 spvpsgcfcgagncshhgngsd-hpgggrefffdhrp-gvfa 112
|:| |::| |::| |::| |::| |::| |::| |::| |::|
QY 74 TPAELRRFDGVQDPRIILMAINGKVFDTKGRKFGPEGPGYGVFA 117
07-Oct-1994
RESULT 15
ENTRY A39402 #type complete
TITLE potassium channel protein IIIA form 1, shaker-type - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change
07-Oct-1994
ACCESSIONS A39402
REFERENCE A39402
#authors McCormack, T.; Vega-Saenz de Miera, E.C.; Rudy, B.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4060
#contents erratum
#accession A39402
#status preliminary
```

```
#molecule_type mRNA
#residues 1-613 #label MCC
#cross-references GB:M34052
SUMMARY #length 613 #molecular-weight 67550 #checksum 8996
Query Match 6.0%; Score 92; DB 14; Length 613;
Best Local Similarity 32.7%; Pred. No. 8.88e+00;
Matches 34; Conservative 14; Mismatches 51; Indels 5; Gaps 5;
Db 12 invggrthetyrstkltp-gtrllalassepgqdcitaaqdklqplpppspprpppl 70
|:| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 16 LESGGLLHEIFTSPLNLLGLCIFLL-YKIVRGDQPAASGDRTTTXXPPL-PRLKRRDF 73
07-Oct-1994
Db 71 spvpsgcfcgagncshhgngsd-hpgggrefffdhrp-gvfa 112
|:| |::| |::| |::| |::| |::| |::| |::| |::|
QY 74 TPAELRRFDGVQDPRIILMAINGKVFDTKGRKFGPEGPGYGVFA 117
07-Oct-1994
Search completed: Thu Nov 6 09:49:24 1997
Job time : 36 secs.
```

\*\*\*\*\*  
[W][A][T][E][R][M][A][N]  
[A][L][G][O][R][I][T][H][M]  
[T][M]  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

on: Thu Jul 23 08:41:51 1998; Maspar time 10.74 Seconds  
748.295 Million cell updates/sec

Tabular output not generated.

Title: >US-08-822-264-1  
Description: (1-220) from US08822264.pap  
Perfect Score: 1541  
Sequence: 1 MAADVATGADPSDLESG.....SISXXYFAKSFYVHVXFT 220

Scoring table: PAM 150  
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir56  
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 44.900; Variance 95.636; scale 0.469

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES |      |       |       | Query Match |        | Description |                       | Pred. No. |  |
|-----------|------|-------|-------|-------------|--------|-------------|-----------------------|-----------|--|
| Alt       | No.  | Score | Match | Length      | DB     | ID          |                       |           |  |
| 1         | 1313 | 85.2  | 194   | 2           | JC5260 |             | progesterone membrane | 8.10e-226 |  |
| 2         | 350  | 22.7  | 152   | 2           | S65181 |             | hypothetical protein  | 5.93e-41  |  |
| 3         | 100  | 6.5   | 327   | 2           | S49796 |             | hypothetical protein  | 9.95e-01  |  |
| 4         | 97   | 6.3   | 154   | 2           | S76882 |             | hypothetical protein  | 2.25e+00  |  |
| 5         | 97   | 6.3   | 492   | 2           | S27160 |             | cytochrome P450 2B12  | 2.25e+00  |  |
| 6         | 96   | 6.2   | 42    | 2           | A34259 |             | cytochrome P450mt4, p | 2.95e+00  |  |
| 7         | 95   | 6.2   | 107   | 1           | JN0811 |             | benzene 1,2-dioxygena | 3.85e+00  |  |
| 8         | 96   | 6.2   | 126   | 2           | J01539 |             | hypothetical 14.5K pr | 2.95e+00  |  |
| 9         | 96   | 6.2   | 243   | 2           | S76182 |             | hypothetical protein  | 2.95e+00  |  |
| 10        | 94   | 6.1   | 116   | 2           | S57106 |             | probable membrane pro | 5.02e+00  |  |
| 11        | 94   | 6.1   | 116   | 2           | S71706 |             | hypothetical protein  | 5.02e+00  |  |
| 12        | 94   | 6.1   | 242   | 2           | B20055 |             | nifv protein - Azotob | 5.02e+00  |  |
| 13        | 94   | 6.1   | 402   | 5           | 1PCAA  |             | Procarboxypeptidase A | 5.02e+00  |  |
| 14        | 93   | 6.0   | 107   | 1           | C36516 |             | toluene dioxygenase ( | 6.53e+00  |  |
| 15        | 92   | 6.0   | 355   | 2           | S64404 |             | hypothetical protein  | 8.47e+00  |  |
| 16        | 93   | 6.0   | 429   | 2           | S50461 |             | mannose-6-phosphate 1 | 6.53e+00  |  |
| 17        | 93   | 6.0   | 430   | 2           | D70193 |             | hypothetical protein  | 6.53e+00  |  |
| 18        | 92   | 6.0   | 469   | 2           | S62702 |             | H+-transporting ATP s | 8.47e+00  |  |
| 19        | 92   | 6.0   | 523   | 2           | D64809 |             | cytochrome d complex  | 8.47e+00  |  |
| 20        | 92   | 6.0   | 581   | 2           | S17150 |             | potassium channel pro | 8.47e+00  |  |
| 21        | 92   | 6.0   | 613   | 2           | A39402 |             | potassium channel pro | 8.47e+00  |  |
| 22        | 92   | 6.0   | 624   | 2           | S22703 |             | voltage-gated potassi | 8.47e+00  |  |
| 23        | 91   | 5.9   | 367   | 2           | S19172 |             | cytochrome P450 2B4 - | 1.10e+01  |  |

|    |    |     |      |   |        |                        |          |
|----|----|-----|------|---|--------|------------------------|----------|
| 24 | 91 | 5.9 | 491  | 1 | O4RTPB | cytochrome P450 2B1 -  | 1.10e+01 |
| 25 | 91 | 5.9 | 491  | 1 | O4RTP2 | cytochrome P450 2B2 -  | 1.10e+01 |
| 26 | 91 | 5.9 | 491  | 2 | A27717 | cytochrome P450 2B5 -  | 1.10e+01 |
| 27 | 90 | 5.8 | 64   | 1 | A47103 | drosocin precursor -   | 1.42e+01 |
| 28 | 89 | 5.8 | 155  | 2 | D69208 | conserved hypothetical | 1.83e+01 |
| 29 | 90 | 5.8 | 327  | 2 | S26647 | phosphoprotein - Pir   | 1.42e+01 |
| 30 | 89 | 5.8 | 411  | 2 | A56610 | corticotropin-releasi  | 1.83e+01 |
| 31 | 89 | 5.8 | 430  | 2 | A56726 | corticosterone recept  | 1.83e+01 |
| 32 | 89 | 5.8 | 431  | 2 | I49279 | sauvagine/corticotrop  | 1.83e+01 |
| 33 | 89 | 5.8 | 431  | 2 | I49149 | CRF receptor - mouse   | 1.83e+01 |
| 34 | 89 | 5.8 | 504  | 2 | S45644 | cytochrome P4502K1 -   | 1.83e+01 |
| 35 | 89 | 5.8 | 784  | 2 | JH0101 | apolipoprotein B-100   | 1.83e+01 |
| 36 | 89 | 5.8 | 836  | 2 | S25218 | papc protein - Escher  | 1.83e+01 |
| 37 | 90 | 5.8 | 1073 | 1 | SV5ECP | carbamoyl-phosphate s  | 1.42e+01 |
| 38 | 88 | 5.7 | 237  | 2 | E55210 | hevi protein - Anabae  | 2.35e+01 |
| 39 | 88 | 5.7 | 375  | 2 | A31134 | actin, macronuclear -  | 2.35e+01 |
| 40 | 88 | 5.7 | 388  | 2 | S26964 | hemoglobin - yeast (P  | 2.35e+01 |
| 41 | 88 | 5.7 | 391  | 2 | S48085 | CAU1 protein - yeast   | 2.35e+01 |
| 42 | 88 | 5.7 | 410  | 1 | B31959 | lysosome-associated m  | 2.35e+01 |
| 43 | 88 | 5.7 | 410  | 2 | JC4317 | lysosome-associated m  | 2.35e+01 |
| 44 | 88 | 5.7 | 419  | 1 | S29127 | carboxypeptidase A (E  | 2.35e+01 |
| 45 | 88 | 5.7 | 1645 | 2 | A37792 | spectrin beta-H chain  | 2.35e+01 |

ALIGNMENTS

RESULT 1

ENTRY JC5260 #type complete

TITLE progesterone membrane binding protein - pig

ORGANISM #formal\_name Sus scrofa domestica #common\_name domestic pig

DATE 25-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 10-Sep-1997

ACCESSIONS JC5260

REFERENCE JC5260

authors Falkenstein, E.; Meyer, C.; Eisen, C.; Scriba, P.C.; Wehling, M.

#journal Biochem. Biophys. Res. Commun. (1996) 229:86-89

#title Full-length cDNA sequence of a progesterone membrane-binding protein from porcine vascular smooth muscle cells.

#contents vascular smooth cell

#accession JC5260

#molecule\_type mRNA

#residues 1-194 #label FAL

#cross-references EMBL:X99714; NID:g1657408; PID:e257707; PID:g1657409

FEATURE 14-42 #domain transmembrane #status predicted #label TMM

SUMMARY 194 #length 194 #molecular-weight 21609 #checksum 6132

|                       |       |  |               |                   |
|-----------------------|-------|--|---------------|-------------------|
| Query Match           | 85.2% | Score 1313;  | DB 2;         | Length 194;       |
| Best Local Similarity | 93.3% | Pred. No. 8.10e-226;                                   |               |                   |
| Matches               | 181;  | Conservative 6;  | Mismatches 6; | Indels 1; Gaps 1; |
| Db                    | 1     | MAADVATGADPSLEGGGLLHEIFTSPNLLGLCFLFYKIVRGDQPAAS-DSDDD  | 59            |                   |
| Qy                    | 1     | MAADVATGADPSLEGGGLLHEIFTSPNLLGLCFLFYKIVRGDQPAASGDRIT   | 60            |                   |
| Db                    | 60    | EPPLPRKRRDFTPAELRRFDGVQDPRLMAINGKVFDTVKRKYGEGYGVFAGRD  | 119           |                   |
| Qy                    | 61    | XPPLPRKRRDFTPAELRRFDGVQDPRLMAINGKVFDTVKRKYGEGYGVFAGRD  | 120           |                   |
| Db                    | 120   | ASRGATCLCKEALKDEYDLSLTPAQOETLNWDQSFTEKYHHVKKLKEGEPTVY  | 179           |                   |
| Qy                    | 121   | ASRGATCLCKEALKDEYDLSLTAAQOETLSQWESQFTFKYHHVKKLKEGEPTVY | 180           |                   |
| Db                    | 180   | SDEEPPKDESARKN 193                                     |               |                   |
| Qy                    | 181   | SDEEPPKDESSRKN 194                                     |               |                   |

RESULT 2

ENTRY S65181 #type complete

TITLE hypothetical protein YPL170w - yeast (Saccharomyces cerevisiae)

```
ALTERNATE_NAMES      hypothetical protein P2515
ORGANISM              #formal_name Saccharomyces cerevisiae
DATE                  10-Dec-1994 #sequence_revision 31-May-1996 #text_change
                        05-Dec-1997
ACCESSIONS            S65181; S69429
REFERENCE              S65134
#authors              Purnelle, B.; Coster, F.; Goffeau, A.
#submission           submitted to the Protein Sequence Database, May 1996
#accession            S65181
#molecule_type       DNA
#residues              1-152 #label PUR
#cross-references      EMBL:Z73526; NID:g1370358; PID:e247050; PID:g1370359;
                        MIPS:YPL170w
#experimental_source   source strain S288C (AB972)
REFERENCE              S69428
#authors              Purnelle, B.; Combiez, S.; Coster, F.; Naveau, F.; Goffeau,
                        A.
#submission           submitted to the EMBL Data Library, March 1996
#description           The sequence of 55 kb on the left arm of yeast
                        identifies 28 open reading frames including 18 unknown
                        among which a new putative serine/threonine protein kinase,
                        a homologue to the human phosphotyrosyl phosphatase
                        activator PTPA and a homologue to the plant pleiotropic
                        regulator PRL1 of PP1 and PP2a phosphatases.
#accession            S69429
#molecule_type       DNA
#residues              1-152 #label PUW
#cross-references      EMBL:X96770; NID:g1403537; PID:e239033; PID:g1403539
GENETICS
#map_position         16L
SUMMARY               #length 152 #molecular-weight 16757 #checksum 5922
Query Match           22.7%; Score 350; DB 2; Length 152;
Best Local Similarity 47.6%; Pred. No. 5.95e-41;
Matches 50; Conservative 21; Mismatches 31; Indels 3; Gaps 2;
Db 43 NFFPRTLSEFNHDDKIFIAIRKGVYDCTGRQFGPSGPTNFAGHDASRGLALNSFD 102
QY 72 DTPAELRRFDGVQDPRILMAINGKVFDTYTKGRKFGPEGPGYVFGAGRASRLATFCLD 131
Db 103 LDVIKDWQDIPDLDTREIDALDEWQHFNKYPICIGTLIPE 147
QY 132 KEALKD-E--YDDLSDLTAAQQTSLSDWESQFTFKYHHVGKLLKE 173
RESULT 3
ENTRY                S49796 #type complete
TITLE                hypothetical protein YIL084c - yeast (Saccharomyces
ORGANISM              cerevisiae)
DATE                  13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change
                        06-Feb-1998
ACCESSIONS            S49796
REFERENCE              S49786
#authors              Connor, R.; Churcher, C.
#submission           submitted to the EMBL Data Library, November 1994
#accession            S49796
#molecule_type       DNA
#residues              1-327 #label CON
#cross-references      GB:Z47047; EMBL:Z46728; NID:g603997; PID:g763262;
                        MIPS:YIL084c
GENETICS
#gene                SGD:SDS3
#cross-references      SGD:S0001346; MIPS:YIL084c
#map_position         9L
SUMMARY               #length 327 #molecular-weight 37625 #checksum 238
Query Match           6.5%; Score 100; DB 2; Length 327;
Best Local Similarity 23.2%; Pred. No. 9.95e-01;
Matches 22; Conservative 28; Mismatches 41; Indels 4; Gaps 4;
Db 74 DLVLRLRLEFYRVSRSIGIEFQEDIEKAEKLIKICKERLYSSIE-OKIKLQER 132
```

```
QY 31 NLLLLGLCLFLYKIVR-GQPAASGDRTTTPPLPRL-KRRDFTPAELRRFDGVQDPR 88
Db 133 LLMQDVANVHYANNYSRPOYOKNTRSHTVSGWDSS 167
QY 89 ILMAI-NGKVFDYTKGRKFGPEGPGYGVFAGRDAS 122
RESULT 4
ENTRY                S76882 #type complete
TITLE                hypothetical protein - Synecchocystis sp. (PCC 6803)
ORGANISM              PCC 6803
#variety              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                        09-Sep-1997
ACCESSIONS            S76882
REFERENCE              S74322
#authors              Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                        Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
                        Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                        Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                        S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                        Yasuda, M.; Tabata, S.
#journal              DNA Res. (1996) 3:109-136
#title                Sequence analysis of the genome of the unicellular
                        cyanobacterium Synecchocystis sp. PCC6803. II. Sequence
                        determination of the entire genome and assignment of
                        potential protein-coding regions.
#cross-references      MIMD:97061201
#accession            S76882
#status               nucleic acid sequence not shown; translation not shown
#molecule_type       DNA
#residues              1-154 #label KAN
#cross-references      EMBL:D90917; NID:g1653836; PID:d1019527; PID:g1653884
#note                 the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
GENETICS
#start_codon          GTG
SUMMARY               #length 154 #molecular-weight 17612 #checksum 2775
Query Match           6.3%; Score 97; DB 2; Length 154;
Best Local Similarity 33.3%; Pred. No. 2.25e+00;
Matches 15; Conservative 13; Mismatches 14; Indels 3; Gaps 3;
Db 88 GMGAGRKIVAACIQK-SLR-EYPNL-DLQISAQAHLETFYAEFGF 129
QY 118 GRDASRLATFCLDKALKDEYDDLSDLTAAQQTSLSDWESQFTF 162
RESULT 5
ENTRY                S27160 #type complete
TITLE                cytochrome P450 2B12 - rat
ORGANISM              #formal_name Rattus norvegicus #common_name Norway rat
DATE                  07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
                        08-Sep-1997
ACCESSIONS            S27160; S18907
REFERENCE              S27160
#authors              Friedberg, T.; Grassow, M.A.; Bartlomowicz-Oesch, B.;
                        Siegem, P.; Arand, M.; Adesnik, M.; Oesch, F.
#journal              Biochem. J. (1992) 287:775-783
#title                Sequence of a novel cytochrome CYP2B cDNA coding for a
                        protein which is expressed in a sebaceous gland, but not in
                        the liver.
#accession            S27160
#molecule_type       mRNA
#residues              1-492 #label FRI
#cross-references      EMBL:X63545; NID:g56049; PID:g56050
#note                 the authors translated the codon CAT for residue 28 as
                        Arg and GAC for residue 83 as Gly
GENETICS
#gene                CYP2B12
CLASSIFICATION         #superfamily cytochrome P450
KEYWORDS               chromoprotein; electron transfer; endoplasmic reticulum;
```





```

##note      the nucleotide sequence was submitted to the EMBL Data
            Library, October 1995
SUMMARY    #length 116 #molecular-weight 13145 #checksum 2626

Query Match        6.1%; Score 94; DB 2; Length 116;
Best Local Similarity 36.4%; Pred.No. 5.02e+00;
Matches 12; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

Db   5  ESKKKNIHAFYPLSPYLFFSSNFGSVHIIFK 37
     : |||: |||: || | :||: ||: ||: ||: ||
QY  188 DESSRKNVAFSGSISXX-YFAKSFVTYHXVF 219

RESULT 12
ENTRY   B32055 #type complete
TITLE   nlfy protein - Azotobacter vinelandii
ORGANISM #formal_name Azotobacter vinelandii
DATE    13-Jul-1989 #sequence_revision 03-Aug-1992 #text_change
       20-Mar-1998

ACCESSIONS B32055
REFERENCE A32055
#authors Jacobson, M.R.; Briglie, K.E.; Bennett, L.T.; Setterquist,
         R.A.; Wilson, M.S.; Cash, V.L.; Beynon, J.; Newton, W.E.;
         Dean, D.R.
#journal J. Bacteriol. (1989) 171:1017-1027
#title Physical and genetic map of the major nif gene cluster from
        Azotobacter vinelandii.
#cross-references MUID:89123097
#accession B32055
#status preliminary
#molecule_type DNA
##residues 1-242 ##label JAC
##cross-references GB:M20568; NID:g758356; PID:g142348
#length 242 #molecular-weight 26702 #checksum 6919

SUMMARY

Query Match        6.1%; Score 94; DB 2; Length 242;
Best Local Similarity 29.7%; Pred.No. 5.02e+00;
Matches 19; Conservative 15; Mismatches 26; Indels 4; Gaps 4;

Db   171 LLADCO-LLYTLSIGPPAAKVVRAGVHPVLRLARPARPAREIV-BEOLKVLATAPPPWLAK 228
QY   34 LUGLCIFLYKIVRGDQQAASDRTTTTYPPLPKLR-RDTFPAELRR-FDCGVQDPRIIM 91

Db   229 AMGA 232
QY   92 AING 95

RESULT 13
ENTRY   lPCAA #type complete
TITLE   Procarboxypeptidase a (EC 3.4.12.2), chain A - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
#note pancreas
REFERENCE A51328
#authors Guasch, A.; Coll, M.; Aviles, F.X.; Huber, R.
#submission submitted to the Brookhaven Protein Data Bank, October 1991
#cross-references PDB:lPCA
REFERENCE TN006613
#authors Guasch, A.; Coll, M.; Aviles, F.X.; Huber, R.
#journal J. Mol. Biol. (1992) 224:141
#title Three dimensional structure of porcine pancreatic
        procarboxypeptidase a. a comparison of the a and b zymogens
        and their determinants for inhibition and activation.
        TN006614
REFERENCE Coll, M.; Guasch, A.; Aviles, F.X.; Huber, R.
#authors ENBO J. (1991) 10:1
#journal Three-dimensional structure of porcine procarboxypeptidase b:
#title a three-dimensional basis of its inactivity.
COMMENT Resolution: 2.0 angstroms
COMMENT Determination: X-ray diffraction
FEATURE 232-255
SUMMARY #disulfide_bonds
        #length 402 #molecular-weight 45126 #checksum 8960

```





328, 399, 408; 534/14, 15; 564/18, 23, 26, 27, 28 [IMAGE AVAILABLE]

14. 5,447,851, Sep. 5, 1995, DNA encoding a chimeric polypeptide comprising the extracellular domain of TNF receptor fused to IgG, vectors, and host cells; Bruce A. Beutler, et al., 435/69.7, 69.5, 320.1, 328, 365; 530/300, 351; 536/23.4 [IMAGE AVAILABLE]

15. 5,248,671, Sep. 28, 1993, Methods and compositions for treatment of cancer using oligonucleotides; Larry J. Smith, 514/44 [IMAGE AVAILABLE]

16. 5,087,617, Feb. 11, 1992, Methods and compositions for treatment of cancer using oligonucleotides; Larry J. Smith, 514/44 [IMAGE AVAILABLE]

=> e goli, surya/in

| E#             | FILE  | FREQUENCY | TERM        |
|----------------|-------|-----------|-------------|
| --             | ----  | -----     | ----        |
| E1             | USPAT | 1         | GOLGER,     |
| LEONID I/IN    |       |           |             |
| E2             | USPAT | 1         | GOLI, MARIA |
| E/IN           |       |           |             |
| E3             | USPAT | 0 -->     | GOLI,       |
| SURYA/IN       |       |           |             |
| E4             | USPAT | 10        | GOLI, SURYA |
| K/IN           |       |           |             |
| E5             | USPAT | 2         | GOLIA,      |
| DOMINICK JR/IN |       |           |             |
| E6             | USPAT | 3         | GOLIA,      |
| KENNETH R/IN   |       |           |             |
| E7             | USPAT | 1         | GOLIA,      |
| MICHEL/IN      |       |           |             |
| E8             | USPAT | 2         | GOLIA,      |
| SALVATORE/IN   |       |           |             |
| E9             | USPAT | 1         | GOLIAN,     |
| TIMOTHY G/IN   |       |           |             |
| E10            | USPAT | 1         | GOLIARD,    |
| FRANCOIS/IN    |       |           |             |
| E11            | USPAT | 1         | GOLIAS,     |
| BERNARD J/IN   |       |           |             |
| E12            | USPAT | 8         | GOLIAS,     |
| JOSEPH H/IN    |       |           |             |

=> s e4

L4 10 "GOLI, SURYA K"/IN

=> d 1-10

1. 5,786,150, Jul. 28, 1998, F.sub.0 ATP synthase subunit; Jennifer L. Hillman, et al., 435/6, 232, 252.3, 320.1; 536/23.2

2. 5,786,148, Jul. 28, 1998, Polynucleotides encoding a novel prostate-specific kallikrein; Olga Bandman, et al., 435/6, 212, 252.3, 320.1; 536/23.2, 24.31 [IMAGE AVAILABLE]

3. 5,776,698, Jul. 7, 1998, Regulation of gene transcription; Jennifer L. Hillman, et al., 435/69.1, 252.3, 320.1, 325; 536/23.1, 23.5 [IMAGE AVAILABLE]

4. 5,763,589, Jun. 9, 1998, Human membrane protein; Jennifer L. Hillman, et al., 536/23.1; 435/320.1; 536/24.3, 24.5 [IMAGE AVAILABLE]

5. 5,763,248, Jun. 9, 1998, CDNA encoding a human ATP synthase Fo subunit (ASYSD); Jennifer L. Hillman, et al., 435/183, 69.1, 252.33, 320.1, 325; 536/23.1, 23.5 [IMAGE AVAILABLE]

6. 5,763,220, Jun. 9, 1998, Human apoptosis-related calcium-binding protein; Jennifer L. Hillman, et al., 435/69.1, 252.3, 254.11, 320.1, 325; 536/23.5 [IMAGE AVAILABLE]

7. 5,756,310, May 26, 1998, CDNA encoding a human phospholemmann-like protein (HPLP); Olga Bandman, et al., 435/69.1, 252.33, 320.1, 325; 536/23.1, 23.5 [IMAGE AVAILABLE]

8. 5,756,299, May 26, 1998, Human carbonyl reductase; Jennifer L. Hillman, et al., 435/6, 199, 252.3, 320.1, 325, 348, 358, 367; 536/23.2 [IMAGE AVAILABLE]

9. 5,734,038, Mar. 31, 1998, Human DBI/ACBP-like protein; Janice Au-Young, et al., 536/23.5; 435/69.3, 252.33, 320.1 [IMAGE AVAILABLE]

10. 5,683,910, Nov. 4, 1997, Human phosphorylase kinase gamma subunit; Olga Bandman, et al., 435/194; 424/94.5; 435/69.1, 193, 252.3, 320.1; 530/350; 536/23.2 [IMAGE AVAILABLE]

=> e falkenstein, eli?/in

E#	FILE	FREQUENCY	TERM
--	----	-----	----
E1	USPAT	1	
FALKENSTEIN, ALBERT J/IN			
E2	USPAT	4	
FALKENSTEIN, DON I/IN			
E3	USPAT	0 -->	
FALKENSTEIN, ELI?/IN			
E4	USPAT	1	
FALKENSTEIN, ELIEZER/IN			
E5	USPAT	11	
FALKENSTEIN, GEORG/IN			
E6	USPAT	1	
FALKENSTEIN, HARRY F JR/IN			
E7	USPAT	1	
FALKENSTEIN, HEINZ P/IN			
E8	USPAT	1	
FALKENSTEIN, HERMANN J/IN			
E9	USPAT	1	
FALKENSTEIN, HERMANN JOSEF/IN			
E10	USPAT	1	
FALKENSTEIN, KENT C/IN			
E11	USPAT	1	
FALKENSTEIN, ROGER/IN			
E12	USPAT	2	
FALKENSTEIN, RUDOLF/IN			

=> e selmin, orn?/in

E#	FILE	FREQUENCY	TERM
--	----	-----	----
E1	USPAT	1	SELMEZCI,
ANDRAS/IN			

E2 USPAT  
FATHI/IN  
E3 USPAT  
ORN?/IN  
E4 USPAT  
PAOLO/IN  
E5 USPAT  
CARLOS A/IN  
E6 USPAT  
ERICH/IN  
E7 USPAT  
ALLEN J/IN  
E8 USPAT  
MARC D/IN  
E9 USPAT  
ARNOLD/IN  
E10 USPAT  
JERRY N/IN  
E11 USPAT  
HAROLD G/IN  
E12 USPAT  
LESTER L/IN

1 SELMI,  
0 --> SELMIN,  
1 SELMIN,  
1 SELMONOSKY,  
4 SELNA,  
5 SELNER,  
3 SELNER,  
1 SELNES,  
1 SELNESS,  
28 SELNICK,  
4 SELNICK,

\*\*\*\*\*  
W P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

March\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
on: Thu Jul 23 09:01:47 1998; MasPar time 790.69 Seconds  
1331.189 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-822-264-2  
Description: (1-788) from US0822264.seq  
Perfect Score: 784  
N.A. Sequence: 1 GCCGCCGAACCCGCGCGCC.....AATATTTAGAAAGTTTGAGC 788  
Comp: CGCGCGCTTGGGGCGCGCG.....TTATAAATCTTCAACATCG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1759237 seqs, 667866413 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est54  
1:em\_est1 2:em\_est3  
Database: genbank-est106  
3:gb\_est1 4:gb\_est10 5:gb\_est11 6:gb\_est12 7:gb\_est13  
8:gb\_est14 9:gb\_est15 10:gb\_est16 11:gb\_est17  
12:gb\_est18 13:gb\_est19 14:gb\_est20 15:gb\_est21  
16:gb\_est22 17:gb\_est23 18:gb\_est24 19:gb\_est25  
20:gb\_est26 21:gb\_est27 22:gb\_est28 23:gb\_est29 24:gb\_est30  
25:gb\_est31 26:gb\_est32

Statistics: Mean 10.822; Variance 2.039; scale 5.307

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	470	59.9	531	10	AA428520 zw47c07.r1 Soares tota	0.00e+00
2	449	57.3	520	21	W58474 zc59b01.r1 Soares para	0.00e+00
3	433	55.2	501	12	AA081900 zn23g04.r1 Stratagene	0.00e+00
4	408	52.0	534	15	AA048529 EST191289 Normalized r	0.00e+00
5	402	51.3	535	24	AA184910 mt62c03.r1 Soares 2NDM	0.00e+00
6	400	51.0	458	23	N52291 yv48a12.s1 Soares feta	0.00e+00
7	387	48.4	501	23	AA023630 mh80f05.r1 Soares mous	0.00e+00
8	383	48.9	508	24	AA288129 vb12e11.r1 Soares mous	0.00e+00
9	381	48.6	528	24	AA274818 vb02b12.r1 Soares mous	0.00e+00
10	380	48.5	512	22	AA117455 mn28g11.r1 Beddington	0.00e+00
11	377	48.1	429	22	AA101294 zn71f03.r1 Stratagene	0.00e+00
12	352	44.9	476	22	AA111285 mc53h02.r1 Life Tech m	0.00e+00
13	347	44.3	390	16	AA166645 zq39b04.r1 Stratagene	0.00e+00

14	336	42.9	438	5	AA389446	mp19e12.r1 Life Tech m	0.00e+00
15	331	42.2	411	22	H95223	Yw59f10.r1 Soares plac	0.00e+00
16	330	42.1	406	22	AA115422	z185e09.r1 Stratagene	0.00e+00
17	329	42.0	369	14	H59281	Yg97g07.r1 Homo sapien	0.00e+00
18	328	41.8	438	18	H23045	Ym51d08.r1 Homo sapien	0.00e+00
19	327	41.7	429	12	AA699865	z181e01.s1 Soares feta	0.00e+00
20	326	41.6	429	9	AA572534	v184d11.r1 Stratagene	0.00e+00
21	325	41.5	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
22	324	41.4	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
23	323	41.3	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
24	322	41.2	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
25	321	41.1	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
26	320	41.0	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
27	319	40.9	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
28	318	40.8	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
29	317	40.7	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
30	316	40.6	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
31	315	40.5	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
32	314	40.4	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
33	313	40.3	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
34	312	40.2	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
35	311	40.1	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
36	310	40.0	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
37	309	39.9	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
38	308	39.8	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
39	307	39.7	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
40	306	39.6	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
41	305	39.5	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
42	304	39.4	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
43	303	39.3	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
44	302	39.2	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00
45	301	39.1	438	23	AA021062	ze67a12.r1 Soares reti	0.00e+00

ALIGNMENTS

RESULT 1  
LOCUS AA428520 531 bp mRNA EST 16-OCT-1997  
DEFINITION zw47c07.r1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone  
773196 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING ATPASE ; ,  
mRNA sequence.  
ACCESSION AA428520  
NID g2112535  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 531)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
TITLE WashU-Merck EST Project 1997  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL: contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 475.  
Location/Qualifiers  
/organism="Homo sapiens"  
/note="Vector: pT73B-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand  
was prepared from mRNA obtained from pooled 8-9 week  
(total) fetus material with a Not I - oligo(dT) primer (5'

TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTTTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo.  
 /db\_xref="taxon:9606"  
 /clone="773196"  
 /clone\_lib="Soares total fetus ND2HF8 9w"  
 /dev\_stage="8-9 weeks"  
 /lab\_host="DH10B"  
 BASE COUNT 99 a 172 c 162 g 95 t 3 others  
 ORIGIN

Query Match 59.98; Score 470; DB 10; Length 531;  
 Best Local Similarity 95.98; Pred. No. 0.00e+00;  
 Matches 514; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

Db 1 ACCCGCGCGCCACTCGCTCGCTCAGAGGAGGAGAAAGTGGCGAGTTCGGGATCCCTGC 60  
 9 ACCCGCGCGCCACTCGCTCGCTCAGAGGAGGAGAAAGTGGCGAGTTCGGGATCCCTGC 68  
 61 CTAGCGCGCGCCAACTTACTCCAGAGATCATGCTGCCGAGGATGTGTGGCGACTGG 120  
 69 CTAGCGCGCGCCAACTTACTCCAGAGATCATGCTGCCGAGGATGTGTGGCGACTGG 128  
 Db 121 CGC-GACCAAGCGATCTGGAGAGCGCGGCTGCTGCATGAGATTTACGTCGCCGCT 179  
 129 CGCGGACCAAGCGATCTGGAGAGCGCGGCTGCTGCATGAGATTTACGTCGCCGCT 188  
 Db 180 CAACCTGCTGCTTGGCTCTGCATCTCTCTCTACAGATCTGCGCGGGACCA 239  
 189 CAACCTGCTGCTTGGCTCTGCATCTCTCTCTACAGATCTGCGCGGGACCA 248  
 Db 240 GCGCGCGCGC-AGCGGAN-CACGACGACGACGA-GCGCGCGCGCTCTGCGCGCGCGCA 416  
 249 GCGCGCGCGCAGCGGACGACGACGACGACGACGCGCGCGCTCTGCGCGCGCGCA 428  
 Db 297 CGCGANCTTACCGCGCGCGAGTCTGCGCGCTTCCAGGCGCTCCAGGACCGCGCATACT 356  
 309 GCGCGACTTACCGCGCGCGAGTCTGCGCGCTTCCAGGCGCTCCAGGACCGCGCATACT 368  
 Db 357 CATGGCCATCAACGGCAAGTGTTCGATGTGACCAAGGCGCAAAATTCACGGCGCGCA 416  
 369 CATGGCCATCAACGGCAAGTGTTCGATGTGACCAAGGCGCAAAATTCACGGCGCGCA 428  
 Db 417 GCGGTANTG-GGGGTCTTTGCTGGAAGAGATGCATCCAGGGCGCTTGCACAAATTTTGCCT 475  
 429 GGGGCGGTATGGGTCTTTGCTGGAAGAGATGCATCCAGGGCGCTTGCACAAATTTTGCCT 488  
 476 GGATAGGAAGCACTGAAGATGAGTACGATGACCTTTCTGACCTACTGCTGCC 531  
 489 GGATAGGAAGCACTGAAGATGAGTACGATGACCTTTCTGACCTACTGCTGCC 544

RESULT 2  
 LOCUS W56474 520 bp mRNA EST 03-JUN-1996  
 DEFINITION zc59b01.r1 Soares parathyroid tumor NDHPA Homo sapiens cDNA clone  
 326569 5', mRNA sequence.  
 ACCESSION W56474  
 NID G1358332  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Euthera; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 520)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 TITLE The WashU-Merck EST Project

JOURNAL COMMENT

Unpublished (1995)

Contact: Wilton RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 451.

FEATURES source

1..520  
 /organism="Homo sapiens"  
 /note="Organ: parathyroid gland; Vector: pT73b  
 (Pharmacia) with a modified polylinker; Site\_1: Not I;  
 Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer  
 [5'-  
 TGTTACCAATCTGAAGTGGAGCGCGCAACAAATTTTTTTTTTTTTTTTTT  
 T-3'], double-stranded cDNA was size selected, ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of a modified pT73  
 vector (Pharmacia). Library went through one round of  
 normalization to a cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
 adenomas was kindly provided by Dr. Stephen Marx, National  
 Institute of Diabetes and Digestive and Kidney Diseases,  
 NIH."  
 /db\_xref="taxon:9606"  
 /clone="326569"  
 /clone\_lib="Soares parathyroid tumor NDHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 mRNA <1..>520  
 BASE COUNT 100 a 160 c 155 g 97 t 8 others  
 ORIGIN

Query Match 57.38; Score 449; DB 21; Length 520;  
 Best Local Similarity 94.58; Pred. No. 0.00e+00;  
 Matches 495; Conservative 0; Mismatches 24; Indels 5; Gaps 5;

Db 1 GAGAAAGTGGCGAGTTCGGATCCCTGCTAGCNCGCCCAACCTTTACTCCAGAGATCA 60  
 QY 41 GAGAAAGTGGCGAGTTCGGATCCCTGCTAGCNCGCCCAACCTTTACTCCAGAGATCA 100  
 Db 61 TGCTGCGGAGGATGTGTGGCGACTGGCGG-GACCCCAAGCGATCTCGAGAGCGCGGC 119  
 QY 101 TGCTGCGGAGGATGTGTGGCGACTGGCGCGCGACCCCAAGCGATCTCGAGAGCGCGGC 160  
 Db 120 TGCTGCGATGAGATTTTCACTGCGCGCTCAACCTGCTGCTGGCTCTGCATCTTCC 179  
 QY 161 TGCTGCGATGAGATTTTCACTGCGCGCTCAACCTGCTGCTGGCTCTGCATCTTCC 220  
 Db 180 TGCTCTACAAGATCTGTGGCGGGGACACCGCGCGCNC-AGCGCGACGACGACGACGACGA 238  
 QY 221 TGCTCTACAAGATCTGTGGCGGGGACACCGCGCGCGCGCGGACGACGACGACGACGA 280  
 Db 239 -GCCNCCCCTCTGCCCGCGCTCAAGCG-GCCANCTTACCCCCCGCGAGCTCGCGGCNT 296  
 QY 281 NGCGCGCCCTCTGCCCGCGCTCAAGCGCGCGACTTCAACCCCGCGAGCTCGCGCGCT 340  
 Db 297 TCGAGCGCTCCAGGACCGCGCATACTATGCGCATCAACGCGCAAGGTGTTCGATGTGA 356  
 QY 341 TCGAGCGCTCCAGGACCGCGCATACTATGCGCATCAACGCGCAAGGTGTTCGATGTGA 400  
 Db 357 CCAAGGCCCAAAATTTACGGGCGCGGACGCGGTANTINGGGGTCTTTGCTGGAGAGATG 416  
 QY 401 CCAAGGCCCAAAATTTACGGGCGCGGCGCGGTATGTTGGGTCTTTGCTGGAGAGATG 460



Db 417 CATCCAGGGGCTTCCACATTTTGCCTGGATATAGGAGCACTGAAGGATGAGTACGATG 476  
 QY 461 CATCCAGGGGCTTCCACATTTTGCCTGGATATAGGAGCACTGAAGGATGAGTACGATG 520  
 Db 477 ACCTTTCTGACCTACATTCGTGCGCCAGAGGAGAGACTCTGAGTGA 520  
 QY 521 ACCTTTCTGACCTCACT-GCTGCGCCAGCAGGAGACTCTGAGTGA 563

RESULT 3  
 LOCUS AA081900 501 bp mRNA EST 23-DEC-1997  
 DEFINITION zn23904.r1 Stragatene neuroepithelium NT2RAMI 937234 Homo sapiens  
 cDNA clone 5' similar to TR:E247050 E247050 CHROMOSOME XVI  
 READING FRAME ORF YPL170W. ;, mRNA sequence.  
 ACCESSION AA081900  
 NID G1623958  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 501)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,  
 Waterston, R., Willamson, A., Wohlmann, P. and Wilson, R.  
 TITLE WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 WARNING: There is evidence that suggests that the 384-well parent  
 plate of this clone contains both human and mouse derived clones.  
 Thus, the origin of this clone is uncertain. This caution should be  
 kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 398.

FEATURES  
 source  
 1. .501  
 /organism="Homo sapiens"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2  
 (Ntera-2/cl.D1) precursor cells induced with Retinoic  
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
 Vector: -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'."  
 /db\_xref="GDB:3926826"  
 /db\_xref="taxon:9606"  
 /clone="548310"  
 /clone\_lib="Stratagene neuroepithelium NT2RAMI 937234"  
 /dev\_stage="Ntera-2/RA+MI neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)."  
 BASE COUNT 96 a 155 c 151 g 92 t 7 others  
 ORIGIN

Query Match 55.2%; Score 433; DB 12; Length 501;  
 Best Local Similarity 95.5%; Pred. No. 0.00e+00;  
 Matches 484; Conservative 0; Mismatches 16; Indels 7; Gaps 6;

Db 1 TGGCGAGTTCGGGATTCCTGCTAGCGGGCCCAACCTTTACTCCAGAGATCATGGCTGC 60  
 QY 48 TGGCGAGTTCGGGATTCCTGCTAGCGGGCCCAACCTTTACTCCAGAGATCATGGCTGC 107  
 Db 61 CGAGAGTGTGTGGGACTTGGCGGCGCCAGCCCAAGCCATCTGNAGCGGGGCTCTGCA 120

QY 108 CGAGATGTGTGGGACTTGGCGGCGGACCCCAAGCGATCTGGAGCCGCGGGCTCTCTCA 167  
 Db 121 TGAGATTTTACAGTCGCGCGCTCAACCTGCTGTGTGCTTGGCCCTCTGCATCTTCTCTCTA 180  
 QY 168 TGAGATTTTACAGTCGCGCGCTCAACCTGCTGTGTGCTTGGCCCTCTGCATCTTCTCTCTA 227  
 Db 181 CAAGATCTGTGGCGGGGACCGCGGCGNC-AGCGGGANCGANGCAGCAGCA-CCCGCC 238  
 QY 228 CAAGATCTGTGGCGGGGACCGCGGCGGCGCCAGCGCGCACAGGACGAGCAGCGCCGCC 287  
 Db 239 CCCTCTGCCCCCGCTCAGCGCGCGCANCTTCAACCCCGCGAGCTGCGG-GCTTGCAGCG 297  
 QY 288 CCCTCTGCCCCCGCTCAGCGCGCGGACTTCAACCCCGCGAGCTGCGGCGCTTGCAGCG 347  
 Db 298 GTCCAGGACCGCGCATACTCATGCCCATCAACGGCAAGGTGTTTCGATGTGACCAAGG 357  
 QY 348 GTCCAGGACCGCGCATACTCATGCCCATCAACGGCAAGGTGTTTCGATGTGACCAAGG 407  
 Db 358 CGCAAAATTTACGGGCGCGGAGCGGTANT--GGGCTTTTCTTGTGGAAGAGATGCATCCAG 415  
 QY 408 CGCAAAATTTACGGGCGCGGAGCGCGGTATGGGTCTTTGCTGGAAGAGATGCATCCAG 467  
 Db 416 GGGCTTTCG-ACATTTTTCCTGGATAAGAGCACTGAAGATGATGATGATGATGATGAT 474  
 QY 468 GGGCTTTCGCAATTTTTCCTGGATAAGAGCACTGAAGATGATGATGATGATGATGAT 526  
 Db 475 CTGACCTAATNGCTGCCGAGAGGAGA 501  
 QY 527 CTGACCTACTGCTGCCAGCAGGAGA 553

RESULT 4  
 LOCUS AA848529 534 bp mRNA EST 06-MAR-1998  
 DEFINITION EST191289 Normalized rat kidney, Bento Soares Rattus sp. CDNA 5'  
 end similar to 25-Dx, mRNA sequence.

ACCESSION AA848529  
 NID G2936069  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 534)  
 AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
 Kerlavage, A.R. and Adams, M.D.

TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat  
 Gene Index  
 JOURNAL Unpublished (1998)  
 COMMENT Other ESTs: EST191288  
 Contact: Lee, NH  
 ATCC

The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@tigr.org  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1. .534  
 /organism="Rattus sp."  
 /note="Organ: kidney; Vector: pT7T3Pac; Site\_1: EcoRI;  
 Site\_2: NotI"  
 /db\_xref="taxon:10118"  
 /clone\_lib="Normalized rat kidney, Bento Soares"

BASE COUNT 107 a 163 c 161 g 102 t 1 others  
 ORIGIN

Query Match 52.0%; Score 408; DB 15; Length 534;  
 Best Local Similarity 89.3%; Pred. No. 0.00e+00;  
 Matches 474; Conservative 0; Mismatches 55; Indels 2; Gaps 2;  
 Db 3 GGAGAGCCGACTTTCGGGATCTCTCATAGCAGGCGCCCAACCTTTGCTCCAGATC 62

QY 40 GGAGAAAGTGGCGAGTTCCGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGATC 99

Db 63 ATGGCTCCGAGAGTGTGGGACTGGCGCGAGCCAGCCAGCGAGCTGGAGGGCGGG 122

QY 100 ATGGCTCCGAGAGTGTGGGACTGGCGCGAGCCAGCCAGCGATCTGGAGAGCGGG 159

Db 123 CTGCTTCAAGAGATTTTACGTGCGCTCTCAACCTGCTGCTTGGCTCTGATCTTC 182

QY 160 CTGCTCATGAGATTTTACGTGCGCTCTCAACCTGCTGCTGCTGCTGCTGCTTTC 219

Db 183 CTGCTCTACAGATGTTTCCGGGGAGCCAGCCGGTCCAGTGGGGACACAGGAGCAG 242

QY 220 CTGCTCTACAGATGTTTCCGGGGAGCCAGCCGGTCCAGTGGGGAGCAGGAGCAG 278

Db 243 GA-GCCGCCCGCTGCGCGGCTCAAGCGCGCTGAGCTTCAACCTGCGCGAGTAAAGCG 301

QY 279 GAGCGCGCGCTGCTGCGCGCTCAAGCGCGCTGAGCTTCAACCTGCGCGAGTAAAGCG 338

Db 302 ATACGATGAGTCCAGGAGCCGCGCTTCTTATGGCCATCAACGGCAAGTGTTCGAGCT 361

339 CTTCGAGCGGCTCCAGGAGCCGCGCTTCTTATGGCCATCAACGGCAAGTGTTCGAGCT 398

362 GACCAAGGCGCAAGTCTTATGGCGGAGGGGCCATACGGGCTTTCGTTGGAAGAGA 421

QY 399 GACCAAGGCGCAAGTCTTACGGCGGAGGGGCCATACGGGCTTTCGTTGGAAGAGA 458

Db 422 TGCATCCAGNGGCTTCCACATTTTCCCTGGACAAAGACACTGAAGATGATGA 481

QY 459 TGCATCCAGNGGCTTCCACATTTTCCCTGGATAGGAAGACACTGAAGATGATGA 518

Db 482 TGACCTTTTGACTCTACTCTGCTGCGGAGGAGACTGAATGACTGGGA 532

QY 519 TGACCTTTTGACTCTACTCTGCTGCGGAGGAGACTCTGAGTACTGGGA 569

RESULT 5 AA184910 535 bp mRNA EST 19-FEB-1997

LOCUS mt62c03.r1 Soares 2NDMT Mus musculus cDNA clone 634468 5' similar

DEFINITION to WP:R07E3.6 CE04722 TRANSLOCATING ATPASE ; mRNA sequence.

ACCESSION AA184910

NID 91768619

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 535)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:386460

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 482.

Location/Qualifiers

1..535

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

FEATURES

source

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

/db\_xref="taxon:10090"

/clone="634468"

/clone\_lib="Soares 2NDMT"

/sex="male"

/tissue\_type="Thymus"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

BASE COUNT 103 a 171 c 159 g 102 t

ORIGIN

Query Match 51.3%; Score 402; DB 24; Length 535;

Best Local Similarity 88.4%; Pred. No. 0.00e+00;

Matches 473; Conservative 0; Mismatches 60; Indels 2; Gaps 2;

Db 3 AAGCGGACTGTTCCGGAGCTCTGCTAGCGCGGCCCAACCTTTGCTCCAGAGATCATGGC 62

QY 45 AAGTGGCGAGTTCGGATCCCTGCTAGCGCGGCCCAACCTTTACTCCAGAGATCATGGC 104

Db 63 TGGCGAGGATGTGGTGGCGACTGGCGC-GACCCGAGCGAGCTAGAGCGGCGGGGCTGCT 121

QY 105 TGGCGAGGATGTGGTGGCGACTGGCGCGGACCAAGCGATCTGGAGAGCGGGGCTGCT 164

Db 122 GCAGGAGATTTTACGCTCTCTCAACCTGCTCTGCTGGGCTCTGCTATCTTCCTGCT 181

QY 165 GCATGAGATTTTACGCTGCGCGCTCAACCTGCTGCTGGGCTCTGCTATCTTCCTGCT 224

Db 182 CTACAAGATGTTTCCGGGAGCAGCGCGTGGCAGTGGGACACAGCAGCAGGAGAA-CC 240

QY 225 CTACAAGATGTTTCCGGGAGCAGCGCGGCGGCCAGCGGCGGCGAGCAGGAGGCC 284

Db 241 ACCCGCGCTGCCCGCTCAAGCGCGCGAACTTCAACCCCTGCGGAGCTGAGGCGCTTCGA 300

QY 285 GCCCGCTCTGCCCGCTCAAGCGCGCGGACTTCAACCCCGCGGAGCTGCGGCGCTTCGA 344

Db 301 TGGCGTCCAGAGCCCGGCTTCTATGCGCATCAAGCGCAAGGTGTTCAGCTGACCAA 360

QY 345 CGCGTCCAGAGCCCGGCTTCTATGCGCATCAAGCGCAAGGTGTTCAGCTGACCAA 404

Db 361 AGGCGCAAGTCTTACGGGCTGAGGGGCACTATGGGCTTTTCCGGAAGAGATGCATC 420

QY 405 AGGCGCAAGTCTTACGGGCTGAGGGGCGGATGGGCTTTTCTGGAAGAGATGCATC 464

Db 421 CAGGGGCTTGGCCACATTTTCCCTGGACAAGAGCACTGAAGGATGAGTATGACGACCT 480

QY 465 CAGGGGCTTGGCCACATTTTCCCTGGATAAGGAGCACTGAAGGATGAGTATGACGACCT 524

Db 481 TTCTGACTCAACCTGTCACAGCAGAGTACCTGAGTACTGGGACTCTCAGTTC 535

QY 525 TTCTGACTCACTGCTGCCAGCAGGAGACTCTGAGTACTGGGAGTCTCAGTTC 579

RESULT 6

LOCUS yv48a12.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone

DEFINITION 245950 3', mRNA sequence.

ACCESSION N52291

NID g1193457

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 458)







Db 185 TCCAGGGGCTTCCACATTTTCCTGGACAAAGACACTGAAGATGATGACGAC 244  
 QY 463 TCCAGGGGCTTCCACATTTTCCTGGATGAAGAGCACTGAAGATGATGACGAC 522  
 Db 245 CATTCTGACTCACCCTGCACAGCAGGAGACCTGAGTACTGGGACTCTCAGTTCACT 304  
 QY 523 CATTCTGACTCACCCTGCACAGCAGGAGACTCTGAGTACTGGGAGTCTCAGTTCACT 582  
 Db 305 TTCAAGTATCATCAGCTGGGAAACCTGCTGAAGAGGGGAGGAGCCTACTGTGTACTCA 364  
 QY 583 TTCAAGTATCATCAGCTGGGAAACCTGCTGAAGAGGGGAGGAGCCTACTGTGTACTCA 642  
 Db 365 GATGATGAAGAACCAAGACAGCAGCTGCGGAAGATGAATGAAGCATTCGGTGA-G 423  
 QY 643 GATGAGGAAGAACCAAGATGAGAGTTCCCGGAAATGCT-TAAGCATTCAGTGAAG 701  
 Db 424 CATATCTATTTTGTATTTTGCAAAATCATTTGTAACATTCAGTCTGTCTTACAACT 483  
 QY 702 TATATCTATNTT-GTATTTTGCAAAATCATTTGTAACAGTCCACTNTGTCTTTAAACAT 760  
 484 GGTGATTTCATATTTAGAA 504  
 761 AGTG-TTCAATATTTAGAA 780

RESULT 11  
 LOCUS AA101294 429 bp mRNA EST 01-DEC-1996  
 DEFINITION zn71f03.r1 Stragene NT2 neuronal precursor 937230 Homo sapiens  
 CDNA clone 563645 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING  
 ATPASE ;, mRNA sequence.

## ACCESSION

AA101294

NID 91648032

KEYWORDS EST

SOURCE human

## ORGANISM

Homo sapiens

Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 429)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,

Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

WashU-Merck EST Project

Unpublished (1995)

## TITLE

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 2058 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham.

Location/Qualifiers

1. .429

/organism="Homo sapiens"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Cloned unidirectionally. Primer: Oligo dt.

Uninduced, exponentially growing neuroepithelial cells

(Ntera-2/cl.D1). Average insert size: 1.0 kb; Uni-ZAP XR

Vector; -5' adaptor sequence: 5' GAATTCGGCAGG 3' -3'

adaptor sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3'

/db\_xref="taxon:9606"

/clone="563645"

/clone\_lib="Stragene NT2 neuronal precursor 937230"

/dev\_stage="Ntera-2 neuroepithelial cells"

/lab\_host="SOLR (kanamycin resistant)"

&lt;1. .429

79 a 136 c 131 g 74 t 9 others

\* BASE COUNT

ORIGIN

Query Match 48.1%; Score 377; DB 22; Length 429;  
 Best Local Similarity 94.0%; Pred. No. 0.00e+00;  
 Matches 404; Conservative 0; Mismatches 24; Indels 2; Gaps 2;

Db 1 AGAAAGTGGCGAGTTCGGATCCCTGCTAGCGGGCCCAACCTTTACTCCAGAGATCAT 60  
 QY 42 AGAAAGTGGCGAGTTCGGATCCCTGCTAGCGGGCCCAACCTTTACTCCAGAGATCAT 101  
 Db 61 GGTGCGCGAGAGTGTGTGGGAGTGGCGGACCCCAAGGATCTGGAGCGCGGGCT 120  
 QY 102 GGTGCGCGAGAGTGTGTGGGAGTGGCGGACCCCAAGGATCTGGAGCGCGGGCT 161  
 Db 121 GCTGATGAGATTTTCACTGCGCGCTCAACCTGCTGCTTGGCTCTGATCTTCC 180  
 QY 162 GCTGATGAGATTTTCACTGCGCGCTCAACCTGCTGCTTGGCTCTGATCTTCC 220  
 Db 181 TGTCTACAAGATCGTGGCGGGGACAGCGCGGCGNCAAGCANGANCAGCAGCAG 240  
 QY 221 TGTCTACAAGATCGTGGCGGGGACAGCGCGGCGGCGGCGGAGCAGCAGCAG 280  
 Db 241 -GCGGCGGCTCTGCGCGGCTCAAGCGCGCNACNTTCAACCGCGGCGGCTGGGGC 299  
 QY 281 NGCGGCGGCTCTGCGCGGCTCAAGCGCGGCGGCTTCAACCGCGGCGGCTGGGGC 340  
 Db 300 TCGAGCGGCTCCAGGACCGCGCATCTCATGCGCATCAACGCGCAAGGTGTTCGATGTA 359  
 QY 341 TCGAGCGGCTCCAGGACCGCGCATCTCATGCGCATCAACGCGCAAGGTGTTCGATGTA 400  
 Db 360 CCAAGGCCGCAAAATTTACGGGCGGCGGCGGNTAATTTGGGCTTTTGGTGGAGAGATG 419  
 QY 401 CCAAGGCCGCAAAATTTACGGGCGGCGGCGGCGGNTAATTTGGGCTTTTGGTGGAGAGATG 460  
 Db 420 CATCCAGGGG 429  
 QY 461 CATCCAGGGG 470

## RESULT 12

## LOCUS

AA11285

476 bp mRNA EST

06-NOV-1996

nc53h02.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus

CDNA clone 557331 5' similar to WP:K07E3.6 CE04722 TRANSLOCATING

ATPASE ;, mRNA sequence.

ACCESSION AA11285

NID 91663163

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Mus.

REFERENCE 1 (bases 1 to 476)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:338123

Seq primer: -28M13 rev1 from Amersham

High quality sequence stop: 455.

Location/Qualifiers



Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 438)  
 Ma, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gaisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Treising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HHMI Mouse EST Project  
 Unpublished (1996)

Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of MedicineP  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:344358

Possible reversed clone: similarity on wrong strand  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 249.  
 Location/Qualifiers  
 1. 438  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /note="vector: pCMV-SPORT2; Site\_1: SalI; Site\_2: NotI;  
 Cloned unidirectionally. Primer: Oligo dt. 8.5dpc  
 embryos. pCMV-SPORT2 vector."  
 /db\_xref="taxon:10090"  
 /clone\_lib="569710"  
 /clone\_lib="Life Tech mouse embryo 8 5dpc 10664019"  
 /dev\_stage="8.5dpc embryos"  
 /lab\_host="DH10B"  
 <1..>438

mrna  
 BASE COUNT 77 a 145 c 137 g 79 t  
 ORIGIN

Query Match 42.9%; Score 336; DB 5; Length 438;  
 Best Local Similarity 90.0%; Pred. No. 0.00e+00;  
 Matches 394; Conservative 0; Mismatches 41; Indels 3; Gaps 3;

Db 2 GGAGGAAGCGGACGTGTTCCGGAGCTCTGCCCTAGCGGGCCCAACCTTTGCTCCAGAGATC 61  
 |||||  
 Qy 40 GGAGAAAGTGGCGAGTTCGGGATCCCTGCTAGCGGGCCCAACCTTTACTCCAGATC 99  
 |||||

62 ATGCGTCCGAGGATGTGGTGGCGACTGGCGCCGACCGAGCGAGCTAGAGGGCGGCGGG 121  
 |||||  
 100 ATGGTCCGAGGATGTGGTGGCGACTGGCGCCGACCCCAAGCGATCTGGAGAGCGGCGGG 159  
 |||||

Db 122 CTGCTGCAGGAGATTTACGCTCTCTCAACCTGCTCCTCTGGGCGCTCTGCATCTTC 181  
 |||||  
 Qy 160 CTGCTGATGAGATTTTACGCTGCGCGCTCAACCTGCTGCTTGGCGCTCTGCATCTTC 219  
 |||||

Db 182 CTGCTCTACAAGATGCTTTCGGGGGACAGCCCGGTGCGAGTGGGGCAACAGCACGACGAC 241  
 |||||

Qy 220 CTGCTCTACAAGATGCTGCGGGGACAGCCGGCGGGCCAGCGGGGCGACAG-GACGACGAC 278  
 |||||

Db 242 GAA-CCACCCCGCTGCCCGCCCTCAAGCG-CGCAAGCTTCAACCCCTCGCGAGCTGAGCGG 299  
 |||||

Qy 279 GANGCGGCGCCCTCTGCGCGCGCTCAAGCGGCGGACCTTCAACCCCGCGAGCTCGGCGG 338  
 |||||

Db 300 TTTCATGGCGCTCAGGACCGCGGATTTCTATGGCCATCAACGCCAAGGTGTTGCGAGT 359  
 |||||

Qy 339 CTTGACGGCGCTCAGGACCGCGGATCTCATGGCCATCAACGCCAAGGTGTTGCGATGT 398  
 |||||

Db 360 GACCAAGGCGCGAAGTTCTACGGGCGCTGAGGGCCCATATGGGTCCTTGGCGGAAGAGA 419  
 |||||

Qy 399 GACCAAGGCGCGAAGTTCTACGGGCCCGAGGGCCGATGGGTCCTTGTGGAAGAGA 458  
 |||||

Db 420 TGCATCCAGGGGCGCTTGC 437



QY 465 CAGGGGCTTCCACATTTTGCCTGGATAAGGAAGCACTGAAGAGTGAAGTACGATGACCT 524  
Db 191 TTCTGACCTCACTGCTGCCAGCNG-AGACTCTGAGTGACTGGGAGTCTCAGTTCACTTT 249  
QY 525 TTCTGACCTCACTGCTGCCAGCAGGAGACTCTGAGTGACTGGGAGTCTCAGTTCACTTT 584  
Db 250 CAAGTATCATCACGTGGGCAAACTGCTGAAGAGGGGAGAGCCCACTGTGTACTCAGA 309  
QY 585 CAAGTATCATCACGTGGGCAAACTGCTGAAGAGGGGAGAGCCCACTGTGTACTCAGA 644  
Db 310 TGAGGAGAACCAAAAGATGAGAGTGCCCGGNAATAATGATTAAAGCATTTCAGTGGAGT 369  
QY 645 TGAGGAGAACCAAAAGATGAGAGTCCCGGNAATAATG-TTAAAGCATT-CAGTGGAGT 702  
Db 370 ATATCTATTTTGTATTTTTCNNAATCATTTGTAAACAGTCC 411  
QY 703 ATATCTATNNT-GTATTTT-GCAAAATCATTTGTAAACAGTCC 742

Job completed: Thu Jul 23 09:49:20 1998  
Job time : 2853 secs.

\*\*\*\*\*  
M P S R L A  
\*\*\*\*\*  
(TW)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
on: Thu Jul 23 08:41:32 1998; MasPar time 900.89 Seconds  
1459.573 Million cell updates/sec  
Tabular output not generated.  
Title: >US-08-822-264-2  
Description: (1-788) from US08822264.seq  
Perfect Score: 784  
N.A. Sequence: 1 GCCGCCGACCCCGCGGCC .....AATATTGAAAGTTTGAGC 788  
Comp: CGCGCGCTTGGCGCGCGG .....TTAATATCTTCRAACTCG

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 457396 seqs, 834335066 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb154  
7:em\_ba 2:em\_htg 3:em\_hum1 4:em\_hum2 5:em\_in 6:em\_om  
7:em\_or 8:em\_ov 9:em\_pat 10:em\_pl 11:em\_ro 12:em\_v1  
Database: genbank106  
13:gb\_ba 14:gb\_htg 15:gb\_in 16:gb\_om 17:gb\_ov 18:gb\_pat  
19:gb\_ph 20:gb\_pl 21:gb\_pr1 22:gb\_pr2 23:gb\_ro 24:gb\_st  
25:gb\_sy 26:gb\_un 27:gb\_v1

Statistics: Mean 10.800; Variance 5.571; scale 1.938  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	700	89.3	1941 22 HSPROGBIN H. sapiens mRNA for put
2	563	71.8	1893 16 SSTERMBP S. scrofa mRNA for ster
3	536	68.4	1786 23 AF042491 Mus musculus putative
4	487	62.1	1885 23 RN063315 Rattus norvegicus 25-D
5	149	19.0	1874 22 HSAJ2030 Homo sapiens mRNA for
6	53	6.8	7218 18 I66494 Sequence 14 from patent
7	40	5.1	215 18 I28278 Sequence 5 from patent
8	35	4.5	7218 18 I66494 Sequence 14 from patent
9	32	4.1	10772 15 AF012089 Drosophila melanogaste
10	31	4.0	215 18 I28278 Sequence 5 from patent
11	30	3.8	10772 15 AF012089 Drosophila melanogaste
12	29	3.7	1540 16 SSBM88 S. scrofa mRNA for BM88
13	28	3.6	354 15 OFU89259 Oxytricha fallax 57kD
14	27	3.4	201 18 A10161 Synthetic DNA for prep
15	27	3.4	201 18 A10158 S. griseus gene for pre

c	16	27	3.4	201 18	A10159	S. griseus gene for pre	2.49e-01
c	17	27	3.4	201 18	A10162	Synthetic DNA for prep	2.49e-01
c	18	27	3.4	565 18	E04076	gDNA encoding envelope	2.49e-01
c	19	26	3.3	827 20	MIPANB5	Podospira anserina mit	8.32e-01
c	20	26	3.3	1739 20	MIPANB6	Podospira anserina mit	8.32e-01
c	21	26	3.3	4061 15	DROEAG	D. melanogaster putativ	8.32e-01
c	22	26	3.3	100314 20	PANMTPACGA	Podospira anserina, ml	8.32e-01
c	23	26	3.3	100314 20	MTFACG	P. anserina complete m	8.32e-01
c	24	25	3.2	201 18	A10159	S. griseus gene for pre	2.71e+00
c	25	25	3.2	201 18	A10162	Synthetic DNA for prep	2.71e+00
c	26	25	3.2	201 18	A10158	S. griseus gene for pre	2.71e+00
c	27	25	3.2	201 18	A10161	Synthetic DNA for prep	2.71e+00
c	28	25	3.2	2209 20	SCYPL170W	S. cerevisiae chromosom	2.71e+00
c	29	25	3.2	2779 13	MXCARD	M. xanthus card gene an	2.71e+00
c	30	25	3.2	4665 13	ECU97665	Erwinia cyripedii mem	2.71e+00
c	31	25	3.2	38734 13	SC10A5	Streptomyces coelicolo	2.71e+00
c	32	25	3.2	48096 15	CET20G5	Caenorhabditis elegans	2.71e+00
c	33	25	3.2	55786 20	SCLACHXVI	S. cerevisiae chromosom	2.71e+00
c	34	25	3.2	152184 14	CEY70G10	Caenorhabditis elegans	2.71e+00
c	35	24	3.1	838 21	HUMHAHEUTR	Human Na/H exchanger g	8.57e+00
c	36	24	3.1	1108 13	CCRFBINA	C. crescentus basal bod	8.57e+00
c	37	24	3.1	1557 16	DOGIDUA03	Dog alpha-L-iduronidas	8.57e+00
c	38	24	3.1	2052 13	STWTEA	S. rimosus tetracycline	8.57e+00
c	39	24	3.1	2423 17	GGU17277	Gallus gallus collapsi	8.57e+00
c	40	24	3.1	2538 27	AC2TKSEA	Avian retrovirus provi	8.57e+00
c	41	24	3.1	6716 13	PSETODG1C	P. putida toluene dioxy	8.57e+00
c	42	24	3.1	20267 22	AB000877	Homo sapiens genomic D	8.57e+00
c	43	24	3.1	22059 16	BTAS1C	B. taurus gene for alph	8.57e+00
c	44	24	3.1	39853 22	AC002522	Homo sapiens Chromosom	8.57e+00
c	45	24	3.1	44896 22	AC004461	Homo sapiens Chromosom	8.57e+00

ALIGNMENTS

1  
RESULT LOCUS HSPROGBIN 1941 bp RNA PRI 30-APR-1997  
DEFINITION H. sapiens mRNA for putative progesterone binding protein.  
ACCESSION Y12711  
NID G2062021  
KEYWORDS progesterone binding protein.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;  
Homo.  
REFERENCE 1 (bases 1 to 1941)  
AUTHORS Falkenstein, E.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1941)  
AUTHORS Falkenstein, E.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-1997) E. Falkenstein, University Heidelberg,  
Institute Clinical Pharmacology Mannheim, Klinikum Mannheim,  
Theodor-Kutzer-Ufer, 68165 Mannheim, FRG  
FEATURES  
Location/Qualifiers  
1. 1941  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="human1"  
/tissue\_type="liver"  
/clone\_lib="psport"  
52. .639  
/codon\_start=1  
/product="putative progesterone binding protein"  
/db\_xref="PID:g2062022"  
/db\_xref="PID:e314174"  
/translation="MAEDVAVATGADPSLESGLLHEIF"SPNLILLIGLIFLYK  
IVRGDPAASGSDDEPPPLPKRRDTPAELRRFGVDQPRITMAINGKVFDTK  
GRKYGPGEYGVAGRDASRGRLATFCLDKALKDEYDDLSLTAAQETLSDWESQF  
TFKHHVHKKLKEGEPTVSDSEEPKDESARKND"  
polya\_site 568 a 409 c 437 g 524 t 3 others  
BASE COUNT 568 a 409 c 437 g 524 t 3 others  
ORIGIN

JOURNAL	Biochem. Biophys. Res. Commun. 229 (1), 86-89 (1996)
MEDLINE	97112407
REFERENCE	2 (bases 1 to 1893)
AUTHORS	Falkenstein, E.
TITLE	Direct Submission
JOURNAL	Submitted (01-AUG-1996) E. Falkenstein, University of Heidelberg, Institute of Clinical Pharmacology, Faculty of Clinical Medicine at Mannheim, Theodor-Kutzer-Ufer, 68135 Mannheim, FRG
COMMENT	Reference: Meyer, C.; J. Biol. Chem. 239; 726-731; 1996 (N-terminal protein-sequence).
FEATURES	Location/Qualifiers
source	1..1893
	/organism="Sus scrofa"
	/sub_species="domestica"
	/db_xref="taxon:9823"
	/tissue_type="aorta"
	/cell_type="smooth muscle cells"
	/clone_lib="Uni-ZAP XR-Vector (Fa. Stratagene)"
	/clone="PgMEF6.1"
	42..626
	/codon_start=1
	/product="steroid membrane binding protein"
	/db_xref="PID:e257707"
	/db_xref="PID:g1657409"
	/translation="MAEDVAATGADPSELEGGLLHEIFTSPLNLLGLCIFIPLYK
	IVRGDPAASDDEPPPLPKRDFTPAELRRFDGVDQPRILMAINGKVFDTYK
	RFYEGPGYGVFAGRADSLATFCLDKALKEVDLSDLTTPAQETLNDWSQFT
	FRYHVHVKLLKEGSEPTVSYDEEKPDESARKND"
BASE COUNT	509 a 447 c 461 g 476 t
ORIGIN	
Query Match	71.8%; Score 563; DB 16; Length 1893;
Best Local Similarity	91.9%; Pred. No. 0.00e+00;
Matches	647; Conservative 0; Mismatches 51; Indels 6; Gaps 5;
Db	21 CCAACCTTTGCTCCAGCGATCATGGCTGCCGAGGATGTCGGCGCTACCGCGCGGACCGG 80
QY	79 CCAACCTTTTACTCCAGAGATCATGGCTGCCGAGGATGTCGGCGACTGGCGCGACCCA 138
Db	81 AGCGAGCTAGAGGCGCGGCTGCTGCATGAGATTTTACGTGCGCGCTCAACCTGCTG 140
QY	139 AGCGATCTGGAGAGCGCGGCTGCTGCATGAGATTTTACGTGCGCGCTCAACCTGCTG 198
Db	141 CTGCTGGCCCTTCGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 200
QY	199 CTGCTTGGCCCTTCGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 258
Db	201 AGCGATAGC--GACGACGACGA--GCCGCCCGCGCTGCCCGCCCTTAAAGCGCGGACTTC 257
QY	259 AGCGGCGACAGGACGACGACGANGCCGCCCTCTGCCCGCCCTCAAGCGCGGCGGACTTC 318
Db	258 ACCCTTGGCGAGCTGCTGCTTCGAGCGGCTCCAGGACCCGCGTATCTACTATGCGCATC 317
QY	319 ACCCGCGGAGCTGCGGCGCTTCGAGCGGCTCCAGGACCCGCGCATCTCATGGCCATC 378
Db	318 AACGGCAAGTGTTCGACGTGACCAAGCGCGCAAGTTCTACGGGCGCGGAGGCGCGTAC 377
QY	379 AACGGCAAGTGTTCGATGTGACCAAGCGCGCAAAATTTCTACGGGCGCGGAGGCGCGTAT 438
Db	378 GGGTCTTTTCTCTGAAGAGACGATCCAGGGGCTGCCACGCTTTTCCCTGGATAAGGAA 437
QY	439 GGGTCTTTTCTCTGAAGAGATGATCCAGGGGCTTCCACATTTTGCCTGGATAAGGAA 498
Db	438 GCCTGAAGACGAGTATGATGACCTTTCTGACCTCACTCTGCCACGAGAGACCTGT 497
QY	499 GCATGAAGATGATGATGATGACCTTTCTGACCTCACTGCTGCCACGAGGACTCTG 558
Db	498 AATGACTGGGACTCTCAGTTTCACTTTCAAGTATCATCAGTGGGCAAACTGCTGAAGAG 557
QY	559 AGTGACTGGAGTCTCAGTTTCACTTTCAAGTATCATCAGTGGGCAAACTGCTGAAGAG 618
Db	558 GGGGAGGAGCGGCGGCTGATGATGAGAGGAGGCGGCAAGATGAGCGGCTCGGAAA 617



SSPSSITWKGKLEGAEBPIVSDDEQKMLLGRVTEAVSGAYLFYFAKSFVTFQS  
VFITW

BASE COUNT 550 a 399 c 427 g 509 t

Query Match 62.18; Score 487; DB 23; Length 1885;  
Best Local Similarity 86.4%; Pred. No. 0.00e+00;  
Matches 658; Conservative 0; Mismatches 90; Indels 14; Gaps 10;

Db 26 TCCTTAGAGGGGAGGAGGAGCGACTGTTTCGG-TCTCTGCAATAACAGGCCCAACCTT 84  
QY 27 TCCTCAGAGGAGGAGAAAGTGGCGAGTTCGGATCCCTGCTAGCGGCGCCCAACCTT 86  
Db 85 TGCTCCAGAGATCATGGCTCCGAGGATGTTGGCGACTGGCCGCCACCCACCGAGCT 144  
QY 87 TACTCCAGAGATCATGGCTCCGAGGATGTTGGCGACTGGCCGCCACCCACCGATCT 146  
Db 145 GGAGGCGGGGGCTGCTTCAAGAGATTTTCAAGCTGCGCTCTCAACCTGCTGCTTGG 204  
QY 147 GGAGAGCGGGGCTGCTGATGAGATTTTCAAGCTGCGCTCAACCTGCTGCTTGG 206  
QY 205 CCTCTCATCTTCTCTACAGATCTTTCGGGGGACAGCCCGTGCAGTGGGA 264  
QY 207 CCTCTCATCTTCTCTACAGATCTTTCGGGGGACAGCCCGTGCAGTGGGA 266  
Db 265 CAACGACGACGAGCA-GCCGCCCGCGCTGCTTCCCTCAAGCCCGTGAATTCACCCCTG 323  
QY 267 CAG-GACGACGAGGAGCCCGCCCTCTGCCCGCTCAAGCGCGGACTTCACCCCG 325  
Db 324 CGAATTAAGCGGATAGATGAGTGGAGTGCAGAGCCCGCGCTTATGCGCCATCAACGCA 383  
QY 326 CCGAGCTGCGGCTTCGACGGGCTCCAGAGCGCGCTCATGCTGCTGCTGCTGCTGCT 385  
Db 384 AGTGTTCGACGTGACCAAGGCGCGAGTTCATGCGCGGAGGACCATACCGGCTCT 443  
QY 386 AGTGTTCGATGTGACCAAGGCGCGAATTTCTACGGGCGCGAGGCGGCTGCTGCT 445  
Db 444 TTCTTGAAGAGATGATCCAGGGGCTTCCACATTTTGGCTGCAAGAAAGCACTGA 503  
QY 446 TTGCTGGAAGAGATGATCCAGGGGCTTCCACATTTTGGCTGCAAGAAAGCACTGA 505  
Db 504 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 563  
QY 506 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 565  
Db 564 GGAATCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 623  
QY 566 GGAATCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 619  
Db 624 CGGAGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 682  
QY 620 GGGAGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 679  
Db 683 GTGACTGAAGCAGTCACTGGA-GCATATCTATTTTGTATTTTCAAAATCATTTGTAAC 741  
QY 680 ATGT-TAAGCATTCAGTGAAGTATATCTATNT-CTATTTGCAAAATCATTTGTAAC 737  
Db 742 ATTCAGTCTGCTTTACACATGAGTGGATTCATATTTAGA 783  
QY 738 AGTCCACTNTGCTTTAAACATAGTG-TTACATATTTAGA 778

RESULT 5 HSAJ2030 1874 bp RNA PRI 27-OCT-1997  
LOCUS Homo sapiens mRNA for putative progesterone binding protein.  
DEFINITION AJ002030  
ACCESSION 92570006  
NID progesterone binding protein.  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1874)

AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
FEATURES  
source

CDS

BASE COUNT 542 a 371 c 459 g 502 t  
ORIGIN

Query Match 19.0%; Score 149; DB 22; Length 1874;  
Best Local Similarity 69.8%; Pred. No. 2.04e-89;  
Matches 263; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Db 284 CCTCTCTGCTTCGATGAAGAGCGGACTTACGCTTGGAGCAGCTGCGCCAGTACGAG 343  
QY 287 CCCCTCTGCGCGCTCAAGCGGCGGACTTACCCCGCGAGCTGCGGCGCTTCGAGC 346  
Db 344 GCTCCCGACCGCGGATCTGCTCGGGTCAATGGGAAAGTCTTCGACGTGACCAAG 403  
QY 347 GCGTCCAGGACCGCGCATCTCATGGCCATCAACGGCAAGTCTTCGATGTGACCAAG 406  
Db 404 GCACCAAGTCTACGCGCGCGGCTCCATATGAATATTTGCTGGTAGGATGCTTCCA 463  
QY 407 GCCCAATCTACGCGCGCGGCGGCTATGGGGTCTTTCGCGAAGAGATGATGCCA 466  
Db 464 GAGGACTGGCCACATTTTGCCTAGATAAAGATGCACCTTAGAGATGAATATGATCTCT 523  
QY 467 GGGGCTTGCACATTTTGCCTGATGAAGAGCACCTGAAGATGAGTACGATGACCTTT 526  
Db 524 CAGATTTGAATGAGTACAATGAGAGTGTTCGAGATGGAAATGCAGTTTAAAGAA 583  
QY 527 CTGACCTCACTGCTGCCAGCAGGAGACTCTGAGTGAAGTGGAGTCTCAGTCTACTTCA 586  
Db 584 AATATGATGATGAGGAGACTCTTAAACAGGAGAGACCATCAGATATACAGATG 643  
QY 587 AGTATCATGCTGGCAAACTGCTGAAGAGGGGGAGGCCCATCTGTGACTCAGATG 646  
Db 644 AAGAAGATACCAAGGAT 660  
QY 647 AGGAGAACCAAAAGAT 663

RESULT 6 I66494 7218 bp DNA PAT 23-DEC-1997  
LOCUS Sequence 14 from patent US 5670367.  
DEFINITION I66494  
ACCESSION 92724471  
NID  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

Gerdes, D.  
Direct Submission  
Submitted (20-OCT-1997) Gerdes D., Institute of Clinical  
Pharmacology Mannheim, University of Heidelberg, Theodor Kutzer  
Ufer 1, Mannheim, 68167, GERMANY  
2 (bases 1 to 1874)  
Gerdes, D.  
Cloning and tissue expression of two putative steroid membrane  
receptors  
Unpublished  
Location/Qualifiers  
1. .1874  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="liver"  
/clone="dg6"  
/clone\_lib="pCMVSPORT"  
7. .678  
/note="putative"  
/codon\_start=1  
/product="progesterone binding protein"  
/db\_xref="PID:e1154367"  
/db\_xref="PID:g2570007"

/translation="MAAGDGVKLTGLGSSSSNDGSGSPDAGAAAGGGWAAAA  
LALLTGGEMLNVALVLYRLVWRGRLGAGAGAGGSESPATSLPRMKRD  
FLEQLROYDGSRNPRILLAVNGKVFDTGSKFYGPAGPYGIFAGDASRLATFCL  
DKDALRDEYDLSLNAVQMESVREWMQFKEYDYVGRLLKPGPEPSEYTDDETDK  
HNKQD"



```
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
join(872, .1000,2310, .2426,6476, .6690,6751, .7707)
/genes="Cp1"
/product="cysteine protease"
872, .7707
/genes="Cp1"
872, .1000
/genes="Cp1"
/number=1
1001, .2309
/genes="Cp1"
2310, .2426
/genes="Cp1"
/number=2
join(2328, .2426,6476, .6690,6751, .7462)
/genes="Cp1"
/codon_start=1
/product="cysteine proteinase-1"
/db_xref="PID:g2305221"
/translation="MTAVLLPFLALLAVAQVSEADVVMEWHHTKLEHRKNVQDET
ERERFKIENKHKHTAKHNQRAEGKVSFKLAVNYADLLHHEFQLMNGPNYTLHK
QLRAADESGKGVTFISPAHVLPKSDVMRTGAVTAVKDGCHGCGWAFSGTGALEGG
HFRKSGVLSLQNVDCSTKYNGCNGGLMDNAPRYIKDNGGIDTESKYPYEAID
DSCHEKGVGADRGFTDIPQDEKKMAEAVATGVPVSAIDASHESFOYSEGVYN
EPQDAQNLDHGLVVGFGTDESGEDYWLKNSWGTWGDGKFIKLRNKENQCGIAS
ASSYPLV"
2427, .6475
/genes="Cp1"
4546, .4553
/genes="Cp1"
/note="Insertion site of P(Casper)(50C)"
/citation=[1]
6476, .6690
/genes="Cp1"
/number=3
6691, .6750
/genes="Cp1"
6751, .7707
/genes="Cp1"
/number=4
join(<8110, .9300,9370, .>9532)
/product="phenylalanyl tRNA synthetase"
join(8110, .9300,9370, .>9532)
/note="potential orf"
/codon_start=1
/product="phenylalanyl tRNA synthetase"
/db_xref="PID:g2305222"
/translation="MLTLRVQAGHLKSTRCLASSAAPKSPSPPOLEVSGSTYA
TDGWTNTPKILSYGVANKHLOTDHPLSIIRQIRVAVFYGAYRNQGNPLFSDQMN
PVTYVQDNLDLIPADHVSQRKSDYIIHQHLLRAHTTAHQVELISGGLDNFLVYG
EYVRDEIDSTVPPFHQADAVRLVKLFERNPGLELEFEETWSGTLPADKILPLHP
SWTKNPSPATRAKMLMEHMKHLVGLTKDLFGPRIKYRWYDTVPFPQSPWELEI
YFKDNVLELGGCIMGHEILQKSGVIGYAGVGLERLAWLFDIPLRLFSWSDS
GFLSQSEKDLNLPVKPISHYPOCTNDLSFWLPDIEVDAGFSPNDFDLVRSVAG
DMVEQISLVDKFKHPTKSSVCFRIVYRHMTLQAEVNEIHKQIASVSDSFVQO
IX"
BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
ORIGIN
Query Match 4.18; Score 32; DB 15; Length 10772;
Best Local Similarity 18.28; Pred. No. 4.17e-04;
Matches 22; Conservative 58; Mismatches 39; Indels 2; Gaps 2;
Db 1678 GARAGTYTKTAAAKRWYAWTMTMKMWTTWKWAM-MKTYRTWMMKMYTSRTTTT 1736
||:|:|: || : : : : : : : : : : : : : : : : : : : : : :
QY 564 GAGAGTCCCGGAAATGTTAAGCATTCAGTGGAGTATCTATNTGTATTGCA 723
||:|:|: || : : : : : : : : : : : : : : : : : : : : : :
Db 1737 SAMMMYTNSTWTKYWAYAWKMKMWTTTARNMAWASWARKWKTSAAAYSAVRKMKWK 1796
||:|:|:|: || : : : : | | : : : : : : : : : : : : : : : :
QY 724 AAATCATTTGTAA-CAGTCCACTNTGCTTTAAACATAGTGTACAATATTAGAAAGT 782
||:|:|:|: || : : : : | | : : : : : : : : : : : : : : : :
Db 1797 W 1797
```

```
QY 783 T 783
RESULT 10
LOCUS 128278 215 bp DNA PAT 30-OCT-1996
DEFINITION Sequence 5 from patent US 5569830.
ACCESSION 128278
NID g1819054
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 215)
AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE Plant inhibitors of fungal polygalacturonases and their use to
control fungal disease
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;
FEATURES Location/Qualifiers
1..215
/organism="unknown"
BASE COUNT 15 a 8 c 25 g 26 t 141 others
ORIGIN
Query Match 4.0%; Score 31; DB 18; Length 215;
Best Local Similarity 14.2%; Pred. No. 1.56e-03;
Matches 25; Conservative 71; Mismatches 78; Indels 2; Gaps 2;
Db 40 TWGVCDDTDTYRVNDSGHNYSSANNYNGNNVGAATHYHTYTNVSGADSKVTDSYN 99
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
Cp 413 TTGGCGCCTTTGGTCACATCGAACACCTTGCCGTGTGATGCCATGAGTATCGCGGGTCC 354
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
Db 100 ASGTSSSNGGTDGNRSGADSYGSSKTAMTSNRRTGKTANNVAVDSRNMGDASVSDKNTKK 159
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
Cp 353 TGGACGCCGTCGAAGCGCCGACGTCGCGGGG-GTGAAGTCGCCGCTTGAGCGGGG 295
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
Db 160 HAKNSADGKVGSKNNGDRNNRYGTGTSNVSNN-CGGNKRDSVSYANNKCCGSSC 214
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
Cp 294 CAGAGGGGGCGGCTCGTCGTCGCTCGCGCTGCGCGCGCGCTGCTGCCCGC 239
||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|: ||:|:|:
RESULT 11
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID g2305220
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 4546 to 4553)
AUTHORS Gray,Y.H., Tanaka,M.M. and Sved,J.A.
TITLE P-element-induced recombination in Drosophila melanogaster: hybrid
element insertion
JOURNAL Genetics 144 (4), 1601-1610 (1996)
MEDLINE 97132596
REFERENCE 2 (bases 1 to 10772)
AUTHORS Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
TITLE Structure of the cysteine proteinase (Cp1) gene of Drosophila
melanogaster and associated mutational effects
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 10772)
AUTHORS Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University
of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES Location/Qualifiers
1..10772
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
```

```

mRNA      join(872, .1000, 2310, .2426, 6476, .6690, 6751, .7707)
/gene="Cp1"
/product="cysteine protease"
872, 7707
/gene="Cp1"
872, .1000
/gene="Cp1"
/number=1
/gene="Cp1"
1001, .2309
/gene="Cp1"
2310, .2426
/gene="Cp1"
/number=2
join(2328, .2426, 6476, .6690, 6751, .7462)
/gene="Cp1"
/codon_start=1
/product="cysteine proteinase-1"
/db_xref="PID:g2305221"
/translation="MRTAVLLPLALLAVAFVADVVMEWHFKLEHRKNYQDET
ERFRUKIFENENKHKAKNOREAGVSKFVAVKADLLHHEFRQLMGFNVTLHK
QLRAADESKGVTFFISPAHVTLPKSDWRTKGAIVAKDQHGCGSCWAFSTGALGQ
HFKSGVLSQNLVDCSTKYGNGCNGGLMDNAPRIKDNKGIDTEKSYFEAID
HSCFNKGVAGTDRGFTDIPQDDEKMAEAVATGVFVSAIDASHESFOFYSEGVN
EPQCDQNLDHGLVYVFGTDESGEDYWLKNSWGTGDKGFKMLRNENQCGIAS
ASSYPLV"
2427, .6475
/gene="Cp1"
4546, .4553
/gene="Cp1"
/notes="insertion site of P(CaSpEr)(50C)"
/citation=[1]
6476, .6690
/gene="Cp1"
/number=3
6691, .6750
/gene="Cp1"
6751, .7707
/gene="Cp1"
/number=4
join(<8110, .9300, 9370, .>9532)
/product="phenylalanyl tRNA synthetase"
join(8110, .9300, 9370, .>9532)
/notes="potential orf"
/codon_start=1
/product="phenylalanyl tRNA synthetase"
/db_xref="PID:g2305222"
/translation="MLTLRVQAGRWLKSTRGLASAAAPKSPSPQLESGSTVA
TDGWTNVTPLKLSYGANKHLOTHPLSIIRQIRVNYFYGAYRNORGNPLFSYDQMN
PVYVQNFEDNLLIPADHVSQRKSDCYIIINQHLRAHTTAHQVELLSGIDNELYVG
EYVRDEIDSTHPVFEHQADARLVTKDLFRNPGLELFEETWSGTFLDKPLILPH
SSWTKPNSPATRAVRLMEHEMKHVLVGLTKDFGPRIKRYWYDTPFPQPSWELEI
YFKDNWLEVLGCGIMRHEILQKSGVHQSIGYAFGVLERLAWLFDIPDLRLFSNDS
GFLSQFSEKDLHLPKYKPSHSYPOCTNDLSFWLPDIEVDAGFSPNDFDLVRSVAG
DMVEQILSLVDKFKHPKTKSSVCFRIVYRHWERTLQAEVNEIHKQIASASVDSFNVQ
IX"

BASE COUNT      2929 a  2357 c  2282 g  3046 t  158 others
ORIGIN

Query Match      3.8%; Score 30; DB 15; Length 10772;
Best Local Similarity 20.0%; Pred.No. 5.73e-03;
Matches 16; Conservative 40; Mismatches 23; Indels 1; Gaps 1;

Db 1702 TMKWMWMTWKMAWKTYRWMM-KMYWTSRTTTTSGAMMMWYTWSTWTKYWAYAMKM 1760
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Cp 786 TCRAACTTCTTAATATGTAACTATGTTTAAAGACANAGTGACTGTACAAATGA 727
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1761 WYTRTWARMAMASWARKWK 1780
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Cp 726 TTTTGCATAACANNATAG 707
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|

RESULT 12      SSBM8      1540 bp      RNA      NAM      28-JUN-1995
LOCUS
```

```

DEFINITION      S.scrofa mRNA for BM88 antigen.
ACCESSION      X82027
NID      9557672
KEYWORDS      BM88 antigen.
SOURCE      pig.
ORGANISM      Sus scrofa
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Suidae; Sus.

REFERENCE      1 (bases 1 to 1540)
AUTHORS      Mamalaki,A., Boutou,E., Hurel,C., Patsavoudi,E., Tzartos,S. and
Matsas,R.
TITLE      The BM88 antigen, a novel neuron-specific molecule, enhances the
JOURNAL      differentiation of mouse neuroblastoma cells
MEDLINE      J. Biol. Chem. 270 (23), 14201-14208 (1995)
FEATURES
source
1..1540
/organism="Sus scrofa"
/db_xref="taxon:9823"
/dev_stage="adult"
/tissue_type="pig brain"
/cclone_lib="pig brain kgt11"
1..1540
/gene="BM88"
1..1540
/gene="BM88"
124..546
/gene="BM88"
/codon_start=1
/product="BM88 antigen"
/db_xref="PID:g557673"
/translation="MESRGRSANSPPKADAKATTEARAPVADVGVPSAKKQEAQAEK
OEPPAAPTLPAAKTSALLDNHSLKPAAPSPDAPKPGDGAEEAEASGGP
GGGPGCPFENLPLLVAGGIAMAAAILGVVFLARKK"

BASE COUNT      274 a  525 c  491 g  250 t
ORIGIN

Query Match      3.7%; Score 29; DB 16; Length 1540;
Best Local Similarity 74.6%; Pred.No. 2.06e-02;
Matches 44; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 1209 TGTGCAATTCCTGGGGCAGCTGTGCTGCCAGCCGCCAGACGGCCCTCCCGCC 1267
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 480 TGTGCAAGGCCCTGGATGATCTCTTCAGCAAGACCCCATACGGCCCTCGGGCC 422
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
LOCUS      OFU89259      354 bp      DNA      INV      14-MAR-1997
DEFINITION      Oxytricha fallax 57kd zinc finger/protein chimera gene, partial
cds.
ACCESSION      U89259
NID      g1881675
KEYWORDS      Oxytricha fallax.
ORGANISM      Oxytricha fallax
Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora;
hypotrichs; Stictrichida; Oxytricha.

REFERENCE      1 (bases 1 to 354)
AUTHORS      Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.
TITLE      A proposed superfamily of transposase genes: transposon-like
JOURNAL      elements in ciliated protozoa and a common 'D35E' motif
MEDLINE      Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)
94134747
AUTHORS      2 (bases 1 to 354)
TITLE      Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
JOURNAL      Selection on the protein-coding genes of the TBEl family of
transposable elements in the ciliates Oxytricha fallax and O.
trifallax
```



```
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 354)
AUTHORS Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah,
School of Med. RM5C334, USA, UT 84132, USA
FEATURES
    source
        1..354
            /organism="Oxytricha fallax"
            /strain="9p1"
            /db_xref="taxon:5944"
            /transposon="TBE1"
            /note="this is a bulk sequence that was generated from a
            PCR product that represents many transposon templates"
            <1..>354
            /codon_start=1
            /product="57kd zinc finger/protein chimera"
            /db_xref="PID:g1881676"
            /translation="HTRDLXKHLKAHKKXEXXXXXLXKLRKAREXXXXXX
            OAXEXXXVXXRNXLXSEXTKIMIKIQYKIPVLAQIDLDTSLOSYLEDSFDKVI
            XDHQYAKVYFPFNYL"
            /transl_table=6
CDS
    106 a 42 c 41 g 54 t 111 others
    1..354
    /organism="Oxytricha fallax"
    /strain="9p1"
    /db_xref="taxon:5944"
    /transposon="TBE1"
    /note="this is a bulk sequence that was generated from a
    PCR product that represents many transposon templates"
    <1..>354
    /codon_start=1
    /product="57kd zinc finger/protein chimera"
    /db_xref="PID:g1881676"
    /translation="HTRDLXKHLKAHKKXEXXXXXLXKLRKAREXXXXXX
    OAXEXXXVXXRNXLXSEXTKIMIKIQYKIPVLAQIDLDTSLOSYLEDSFDKVI
    XDHQYAKVYFPFNYL"
    /transl_table=6
BASE COUNT 106 a 42 c 41 g 54 t 111 others
ORIGIN
    1..354
    3 6%; Score 28; DB 15; Length 354;
    Best Local Similarity 26.6%; Pred. No. 7.24e-02;
    Matches 37; Conservative 50; Mismatches 50; Indels 2; Gaps 2;
    Db 30 AAAAGCYCAYARAAGADWTAGARTTGARRAKHAWGAMRHSARYTRAAGYTDCTCAA 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 650 AAGAACCAAAAGATGAGAGTTCGCCGAAAATGTTAAAGCAATTCAGTGAAGTATATCTA 709
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 90 HAAAGAAR-CCDAGRGARYHYHDWARGWG-VDGWCYKAYAAACGACWAGMGARWYA 147
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 710 TNNGTATTTTGCAAAATCATTTGTAACAGTCCACTNTGTCTTTAAACATAGTGTACA 769
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 148 SWRTRYGTAARGAWDKRC 166
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 770 ATATTAGAAAGTTTGAGC 788
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    RESULT 14
    LOCUS A10161 201 bp DNA PAT 01-DEC-1993
    DEFINITION Synthetic DNA for preprotease leader & prochymosin.
    ACCESSION A10161
    NID g490655
    KEYWORDS
        unidentified.
        unclassified.
        unclassified.
    SANISM
    REFERENCE
    1 (bases 1 to 201)
    AUTHORS Garvin,R.T. and James,E.
    TITLE Production of active proteins containing cystine residues
    JOURNAL Patent: EP 0222279-A 5 20-MAY-1987;
    Cangene Corporation
    FEATURES
        Location/Qualifiers
        source
            1..201
            /organism="Artificial sequences"
            1..>201
            /codon_start=1
            /product="preprotease leader & prochymosin"
            /db_xref="PID:g490656"
            /translation="MPHSPVSPAEPQPGRPVSVRRLLGGAAVLGALALSASP
            LTAQAAVRAARETRIPLYKGK"
            LTAQAAVRAARETRIPLYKGK"
BASE COUNT 16 a 66 c 51 g 22 t 46 others
ORIGIN
    1..201
    3 4%; Score 27; DB 18; Length 201;
    Best Local Similarity 36.9%; Pred. No. 2.49e-01;
    Matches 31; Conservative 24; Mismatches 29; Indels 0; Gaps 0;
    Db 23 CSGCGAGTCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTSC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 272 CGACGACGANGCGCCCTCTGCCCCGCTCAAGCGCGCGACTTCACCCCGCGGAGC 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 83 TSGAGGGSGGSGCGSGCTSG 106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 332 TGGCGGCTTCGACGGCGTCCAGG 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Search completed: Thu Jul 23 09:01:28 1998
    Job time : 1196 secs.
```

\*\*\*\*\*

MPERCH\_PP

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mperch\_pp protein - protein database search, using Smith-Waterman algorithm

on: Thu Jul 23 08:43:12 1998; Mapar time 12.99 Seconds

Module output not generated. 713,290 Million cell updates/sec

Title: >US-08-822-264-1  
Description: (1-220) from US08822264.pep  
Perfect Score: 1541  
Sequence: 1 MAEDVATGADPSDLESGG.....SISXXYFAKSFVTVHXVKT 220

Scoring table: PAM 150  
Gap 11

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprembls  
1:sp\_fungi 2:sp\_human 3:sp\_invertebrate 4:sp\_mammal  
5:sp\_mhc 6:sp\_organelle 7:sp\_phage 8:sp\_plant  
9:sp\_bacteria 10:sp\_rodent 11:sp\_virus 12:sp\_vertebrate  
13:sp\_unclassified

Statistics: Mean 44.865; Variance 77.827; scale 0.576

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1349	87.5	195	2	000264 PUTATIVE PROGESTERONE	2.00e-286
2	1313	85.2	194	4	Q95250 STEROID MEMBRANE BINDI	1.05e-277
3	1219	79.1	223	10	P70580 25-DX.	5.70e-255
4	721	46.8	223	2	O15173 PROGESTERONE BINDING	5.97e-136
5	416	27.0	964	3	O10461 HYPOTHETICAL 106.7 KD	1.65e-65
6	367	23.8	166	1	O13995 HYPOTHETICAL 18.9 KD P	1.38e-54
7	350	22.7	152	1	Q12091 CHROMOSOME XVI READING	7.70e-51
8	174	11.3	797	8	O23350 HYPOTHETICAL 88.9 KD P	2.94e-14
9	106	6.9	1377	8	O15377 CARBOXYPEPTIDASE D PRE	1.73e-02
10	104	6.7	2408	2	Q92566 MYELOBLAST KIAA0279 (F	3.46e-02
11	102	6.6	398	9	P77842 RNA POLYMERASE SIGMA F	6.85e-02
12	98	6.4	284	1	O13848 HYPOTHETICAL 31.6 KD P	2.61e-01
13	98	6.4	918	8	Q33307 HYPOTHETICAL 103.5 KD	2.61e-01
14	96	6.2	126	9	O23452 14.5 KD PROTEIN IN MOB	5.03e-01
15	96	6.2	243	9	P74347 GALACTOSYL-1-PROSPHATE	5.03e-01
16	95	6.2	328	3	Q27759 CATHESPIN L PRECURSOR	6.95e-01
17	95	6.2	749	3	O17768 CODED FOR BY C. ELEGAN	6.95e-01
18	94	6.1	521	9	O05191 SUBUNIT 1 OF CYTOCHROM	9.59e-01
19	94	6.1	813	10	P70625 ZONULA OCCLUDENS 2 PRO	9.59e-01
20	92	6.0	469	6	Q36517 ATPASE ALPHA SUBUNIT (	1.81e+00

21	92	6.0	514	9	Q44691 ANTHRANILATE SYNTHASE	1.81e+00
22	92	6.0	624	10	Q63735 VOLTAGE-GATED POTASSIU	1.81e+00
23	92	6.0	699	6	O19855 NADH DEHYDROGENASE-LIK	1.81e+00
24	92	6.0	878	8	O04624 SIMILARITY TO MEMBRANE	1.81e+00
25	91	5.9	491	10	Q64582 CYTOCHROME P450 (EC 1.	2.48e+00
26	91	5.9	491	10	Q64584 CYTOCHROME P450 B (EC	2.48e+00
27	91	5.9	491	4	Q29516 CYTOCHROME P450 (P450I	2.48e+00
28	91	5.9	655	3	Q09536 HYPOTHETICAL 751.1 KD	2.48e+00
29	90	5.8	342	2	O15491 HAN11.	3.38e+00
30	90	5.8	350	9	Q47490 OUTER MEMBRANE PROTEIN	3.38e+00
31	90	5.8	413	10	O35141 MAXP1.	3.38e+00
32	90	5.8	699	6	O19851 NADH DEHYDROGENASE-LIK	3.38e+00
33	90	5.8	699	6	O19848 NADH DEHYDROGENASE-LIK	3.38e+00
34	90	5.8	699	6	O21788 NADH DEHYDROGENASE-LIK	3.38e+00
35	90	5.8	699	6	O19844 NADH DEHYDROGENASE-LIK	3.38e+00
36	90	5.8	699	6	O19853 NADH DEHYDROGENASE-LIK	3.38e+00
37	90	5.8	699	6	O19847 NADH DEHYDROGENASE-LIK	3.38e+00
38	90	5.8	699	6	O19852 NADH DEHYDROGENASE-LIK	3.38e+00
39	90	5.8	699	6	O19845 NADH DEHYDROGENASE-LIK	3.38e+00
40	90	5.8	699	6	O19846 NADH DEHYDROGENASE-LIK	3.38e+00
41	90	5.8	699	6	O19854 NADH DEHYDROGENASE-LIK	3.38e+00
42	90	5.8	699	6	O19850 NADH DEHYDROGENASE-LIK	3.38e+00
43	90	5.8	699	6	O19856 NADH DEHYDROGENASE-LIK	3.38e+00
44	90	5.8	699	6	O19849 NADH DEHYDROGENASE-LIK	3.38e+00
45	90	5.8	1306	9	P77070 ESPC.	3.38e+00

ALIGNMENTS

RESULT 1  
ID O00264 PRELIMINARY; PRT; 195 AA.  
AC O00264;  
DT 01-JUL-1997 (TREMREL. 04, CREATED)  
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JUL-1997 (TREMREL. 04, LAST ANNOTATION UPDATE)  
DE PUTATIVE PROGESTERONE BINDING PROTEIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA FALKENSTEIN E.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Y12711; E314174; --  
SQ SEQUENCE 195 AA; 21671 MW; DE192BEF CRC32;

Query Match 87.5%; Score 1349; DB 2; Length 195;  
Best Local Similarity 96.9%; Pred. No. 2.00e-286;  
Matches 188; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db	1	MAEDVATGADPSDLESGGLHEIFTSPLNLLGLCFLFYKIVRGDQPAASGSDDD	60
Qy	1	MAEDVATGADPSDLESGGLHEIFTSPLNLLGLCFLFYKIVRGDQPAASGDRTT	60
Db	61	EPPLPRKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPGEYVAGRD	120
Qy	61	XPPLPRKRRDFTPAELRRFDGVQDPRILMAINGKVFDTGKRFYGPGEYVAGRD	120
Db	121	ASRGLATCLDKEALKDEYDLSLTAQOETLSWESQFTFKYHVGLKKEGEPTVY	180
Qy	121	ASRGLATCLDKEALKDEYDLSLTAQOETLSWESQFTFKYHVGLKKEGEPTVY	180
Db	181	SDEEPKDESARKN 194	
Qy	181	SDEEPKDESSRKN 194	

RESULT 2  
ID Q95250 PRELIMINARY; PRT; 194 AA.  
AC Q95250;  
DT 01-FEB-1997 (TREMREL. 02, CREATED)  
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)





QY 78 LRFDGQDPRLMAINGKVFDTK-GRFYGEGPYGVFAGRDASRGLATFCLDKALK 136  
Db 219 DOFTSGEPALDEVEPRALIEWIR-N-KFVLSGNL-HGGSVYVAFDDSPSPHKTGI 275  
QY 137 DEYD--DLSLTAQAQE-TLSDWESQFTFKYHHVGLKKEGEPTVYSDEEKPDESSRK 193  
Db 276 YKTSDEDEV-FYLAAYASNHPIMKT 301  
QY 194 NVKAFSGSISXXYFAKSFVTVHXVFKT 220

RESULT 10  
ID Q92566 PRELIMINARY; PRT; 2408 AA.  
AC Q92566;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE MYELOBLAST KIA0279 (FRAGMENT).  
GN KIA0279.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
EUTHERIA; PRIMATES.  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE; 97191544.  
RA NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,  
RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
RL DNA RES. 3:321-329(1996).  
DR EMBL; D87469; D1014097; -.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
FT NON\_TER 1  
SQ SEQUENCE 2408 AA; 261739 MW; CDBA2001 CRC32;

Query Match 6.7%; Score 104; DB 2; Length 2408;  
Best Local Similarity 25.7%; Pred. No. 3.46e-02;  
Matches 19; Conservative 15; Mismatches 36; Indels 4; Gaps 3;

Db 352 ESTSGIVRTLRLDRNQAQYVLRAVADKGMPPARTPMVTVLVDNDNPPVFEODE- 410  
QY 56 DRTTYPPLPRKRDFTPAELRRF--D-GVQDPRLMAINGKVFDTKGRFYGEGP 112

Db 411 FDFVFEENSPIGLA 424  
QY 113 YGVFAGDASRGLA 126

RESULT 11  
ID P77842 PRELIMINARY; PRT; 398 AA.  
AC P77842;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE RNA POLYMERASE SIGMA FACTOR.  
GN SIGD.  
OS CHLOROFLEXUS AURANTIACUS.  
OC PROKARYOTA; GRACILICUTES; ANOXYPHOTOBACTERIA; GREEN BACTERIA;  
OC CHLOROFLEXACEAE.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-J10FL;  
RA GRUBER T.M., BRYANT D.A.;  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
CC THEN IS RELEASED.  
DR EMBL; U67722; G1519642; -.  
DR PROSITE; PS00715; SIGMA70\_1; 1.  
DR PROSITE; PS00716; SIGMA70\_2; 1.  
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;  
KW DNA-BINDING.  
\*SQ SEQUENCE 398 AA; 44238 MW; 8FC0A71C CRC32;

Query Match 6.6%; Score 102; DB 9; Length 398;  
Best Local Similarity 36.2%; Pred. No. 6.85e-02;  
Matches 21; Conservative 12; Mismatches 22; Indels 3; Gaps 3;  
Db 315 ADDAAAGSIVRQITAAALDOLSERERRVLELR-YGLVDGQPTLLEEV-GKAFGVTRER 370  
QY 48 GDQPAASGDRTTTTPPLPRKLRDFTPAELRRDGVQ-DPRILMAINGKVFDTKGR 104

RESULT 12  
ID O13848 PRELIMINARY; PRT; 284 AA.  
AC O13848;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 31.6 KD PROTEIN.  
GN SPAC19G12.10.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYETES.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA OLIVER K., HARRIS D.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z97209; E325344; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 284 AA; 31569 MW; B6078AE2 CRC32;

Query Match 6.4%; Score 98; DB 1; Length 284;  
Best Local Similarity 28.9%; Pred. No. 2.61e-01;  
Matches 13; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

Db 175 AAKPLVFCOSKGIIVGYPGLSPVRDAQGPVAFETKSLSKYH 219  
QY 121 ASRGLATFCLDKALKDEYDDLTLTAQAQETLSDWESQFTFKYH 165

RESULT 13  
ID O23307 PRELIMINARY; PRT; 918 AA.  
AC O23307;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 103.5 KD PROTEIN.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
OC CAPPARALES; CRUCIFERAE.  
RN [1]  
RN SEQUENCE FROM N.A.  
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,  
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A., JONES J.,  
RA PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W., SCHUELLER C.,  
RA CHALWATZIS N.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RN SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z97336; E326895; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 918 AA; 103468 MW; DA52B1F9 CRC32;

Query Match 6.4%; Score 98; DB 8; Length 918;  
Best Local Similarity 23.1%; Pred. No. 2.61e-01;  
Matches 18; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 97 RSGDVGSGSSSTMEKIVEKLKKYGFVDEQFQDKVEQERIEKSSVEERFVEERRG 156

QY 47 RGQPAASGORTTTPPLPKRDETPAELRRFDGV-QDPRILMAINGKVEDVTKGRK 105  
 Db 157 GFSESPGVYGGNDEVK 174  
 QY 106 FYGPEPGYVFGARDASR 123

Search completed: Thu Jul 23 08:43:55 1998  
 Job time : 43 secs.

RESULT 14  
 ID Q03452 PRELIMINARY; PRT; 126 AA.  
 AC Q03452;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE 14.5 KD PROTEIN IN MOBA 3' REGION (ORF2).  
 OS SALMONELLA TYPHIMURIUM.  
 OG PLASMID NTP16.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.

[1]  
 SEQUENCE FROM N.A.  
 TRANSPOSON-IN4353;  
 RX MEDLINE; 92383313.  
 RA CANNON P.M., STRIKE P.;  
 RL PLASMID 27:220-230(1992).  
 DR EMBL; L05392; G150421; -.  
 KW PLASMID.  
 SQ SEQUENCE 126 AA; 14527 MW; 95FA1703 CRC32;

Query Match 6.2%; Score 96; DB 9; Length 126;  
 Best Local Similarity 31.7%; Pred. No. 5.03e-01;  
 Matches 13; Conservative 16; Mismatches 10; Indels 2; Gaps 2;  
 Db 60 GKDRY-VYTSFNGEK-FSSYTLNKVTKTDEYNLSLSAS 98  
 QY 108 GPEPGYVFGARDASRGLATFCLDKALKDEYDLSLTAA 148

RESULT 15  
 ID P74347 PRELIMINARY; PRT; 243 AA.  
 AC P74347;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
 DE GALACTOSYL-1-PHOSPHATE TRANSFERASE.  
 GN RFBP.  
 OS SYNECHOCYSTIS SP.  
 EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA TABATA S.;  
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,  
 RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,  
 RA YAMADA M., YASUDA M., TABATA S.;  
 RL DNA RES. 3:109-136(1996).  
 DR EMBL; D90914; G1653528; -.  
 KW TRANSFERASE.  
 SQ SEQUENCE 243 AA; 27788 MW; 46CF5B19 CRC32;

Query Match 6.2%; Score 96; DB 9; Length 243;  
 Best Local Similarity 25.0%; Pred. No. 5.03e-01;  
 Matches 12; Conservative 16; Mismatches 19; Indels 1; Gaps 1;  
 Db 2 TANSSPISVKALRALMRRCFFYTVSPRRYQGSRSSSLTCTVAKRVEDI 49  
 QY 54 SGRDRTTTPPLPKRDETPA-ELRRFDGVQDPRILMAINGKVEDV 100